CONSISTENCY OF FEIGNING MEASURE SCORES OVER TIME:

A STUDY OF SIMS SCORES ACROSS FIVE ADMINISTRATIONS

A Dissertation

Presented to

The Faculty of the Department of Psychology and Philosophy

Sam Houston State University

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

---

by

Jessica Rae Hart

August, 2020

CONSISTENCY OF FEIGNING MEASURE SCORES OVER TIME:

A STUDY OF SIMS SCORES ACROSS FIVE ADMINISTRATIONS

by

Jessica Rae Hart

APPROVED:

Marcus T. Boccaccini, Ph.D.
Dissertation Chair

David Nelson, Ph.D.
Committee Member

Jaime L. Anderson, Ph.D.
Committee Member

John F. Edens, Ph.D.
Committee Member

Abbey Zink, Ph.D.
Dean, College of Humanities and Social Sciences

ABSTRACT

Hart, Jessica R., *Consistency of feigning measure scores over time: A study of SIMS scores across five administrations*. Doctor of Philosophy (Clinical Psychology), August, 2020, Sam Houston State University, Huntsville, Texas.

Feigning of mental health and neurocognitive symptoms is a serious concern, particularly in forensic settings. While researchers have developed several assessment tools to detect feigning, little research exists regarding the pattern of scores on these measures over time. This dissertation involved secondary data analyses using the publicly available dataset from the study entitled, "Evaluation of the Psychological Effects of Administrative Segregation in Colorado, 2007-2010" (O'Keefe & Klebe, 2014). Participants were 270 adult male inmates who were administered the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) along with measures of psychopathology at five time points (baseline and 3, 6, 9, and 12 month follow-ups). This dissertation sought to investigate patterns of scores on the SIMS across the five administrations in order to examine the consistency of inmates' reporting over time. Additional research questions included comparing two different recommended cut-scores for the SIMS among inmates with and without mental illness and investigating other factors that may influence evaluees' SIMS scores over time. At all time points, inmates with mental illness received significantly higher mean SIMS Total Scores compared to inmates without mental illness. Indeed, 75.5% of inmates with mental illness scored in the feigning range at least once, compared to 35.7% for inmates without mental illness. Test-retest reliability for the SIMS Total Score at all combination of time points revealed moderate to high reliability (*r* ranging from .61 to .83). McNemar's tests indicated similar proportions of participants scored in the feigning range at all

combinations of time points, with the exception of the Time 1 and Time 5 comparison. Repeated measures MANOVAs were used to compare subgroups of inmates (i.e., Never, Sometimes, or Always Feigning) on measures of psychopathology (i.e., the PAS and BPRS). Results revealed inmates in the Always Feigning group tended to receive higher mean scores on the PAS and BPRS compared to inmates in the Sometimes Feigning Group, and both groups scored higher than inmates in the Never Feigning group. Scores on these other measures also tended to decrease from the baseline administration to later time points. Regression models revealed that changes in scores on the PAS and BPRS predicted changes in scores on the SIMS; 25.4% of the variance in the model was explained by these two predictors, with the PAS having the stronger influence. Neither housing placement nor mental health needs significantly predicted changes in SIMS scores; however, a moderate-strength correlation suggested regression to the mean explained some of the change in SIMS scores over time.
KEY WORDS: Feigning, Malingering, SIMS, Structured Inventory of Malingered Symptomatology, Consistency, Cut scores

DEDICATION

To my parents, Carolyn and Steve, for modeling the values of hard work and higher education, for providing opportunities and space to discover my love of learning, for proving it was possible to balance a family and a doctoral program, and for passing down "nuggets" of wisdom for life.

To my husband, Bryan, for moving across the state so I could attend graduate school, for being both a source of encouragement and a much-needed distraction, and for bringing me cupcakes when I stayed up late writing.

To my son, Ender, for providing reasons to spend time outdoors and for reminding me to put away the to-do list occasionally and just enjoy life.

To all of the phenomenal teachers, coaches, professors, and supervisors who have mentored me, for inspiring perseverance and continual growth.

Thank you.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

A primary concern for mental health treatment providers and evaluators is the provision of appropriate treatment to the many individuals who need it, a task that can be difficult given the limited time and resources available in most settings. Thus, it is important to be able to differentiate between individuals truly in need of services and those who may be exaggerating or feigning their symptoms. The need to distinguish between these groups is particularly salient in forensic settings, where an evaluator's opinion regarding symptoms of mental illness can influence the outcome of criminal and civil cases. Researchers have developed several assessment tools for detecting feigned psychopathology, and meta-analyses provide strong support for the ability of scores on many of these measures to detect feigning (Green & Rosenfeld, 2011; Hawes & Boccaccini, 2009; Rogers, Sewell, Martin, & Vitacco, 2003).

Most feigning measure studies report findings based on a single administration of the instrument, an understandable focus given that clinicians are typically interested in detecting feigning at one specific point in time. However, studying patterns of scores across multiple administrations of the same measure has the potential to provide useful information about both the psychometric properties of these measures and how consistent evaluees tend to be in their response styles. For example, based on the current research literature, we do not know how common it is for evaluees to consistently score in the feigning range across multiple administrations of the same measure, or for them to vacillate between the honest and feigning ranges. If evaluees do fluctuate in and out of feigning ranges, it is unknown whether these fluctuations are best explained by corresponding changes in psychopathology symptoms, by changes in response style, by measurement error, or perhaps based on evaluee characteristics. The purpose of this dissertation was to study patterns of scores over multiple administrations of the same feigning screener across a 12-month period, thereby contributing to our knowledge of the consistency of self-reported feigning across time.

The Diagnostic and Statistical Manual of Mental Disorders- Fifth Edition (DSM-5; American Psychiatric Association [APA], 2013) defines malingering as “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs” (APA, 2013, p. 726). A related term, which must be distinguished from malingering, is feigning: “the deliberate fabrication or gross exaggeration of psychological or physical symptoms without any assumptions about its goals” (Rogers, 2018a, p. 6). Thus, assessments developed to evaluate false or exaggerated symptoms can be used to evaluate feigning, but without being able to assess the individual’s specific motivations for feigning, they cannot establish malingering (Melton, Petrila, Poythress, Slobogin, Otto, Mossman, & Condie, 2018; Rogers, 2018a). For this reason, the term feigning will be used throughout the current study, rather than malingering.

It is important to note that the presence of feigning does not eliminate the possibility of genuine symptoms (Melton et al., 2018; Rogers, 2018a). Rather, the term feigning may be used to describe the response style of an individual who is reporting more symptoms, or more severe symptoms, than would be objectively observed. Resnick, West, and Wooley (2018) discussed three categories of malingering: Pure malingering refers to feigning of symptoms that the person does not experience at all. Partial malingering refers to the actions of an individual who experiences genuine symptoms but who purposefully exaggerates them. Finally, false imputation refers to the situation in which an individual experiences genuine symptoms but consciously attributes them to the wrong source. Partial malingering occurs most commonly of these three response styles, while pure malingering is more rare.

Symptom feigning can take numerous forms, including thought disorder or psychosis, anxiety or depression, posttraumatic stress symptoms, medical or pain symptoms, cognitive or neuropsychological impairment, or even impairment of specific psycholegal abilities such as competency to stand trial (Bender, 2018; Granacher & Berry, 2018; Lezak, Howieson, Bigler, & Tranel, 2012; Melton et al., 2018; Resnick & Knoll, 2018; Resnick, West, & Wooley, 2018; Rogers, 2018a; Soliman & Resnick, 2010). Individuals choosing to feign may do so within just one of these domains or may combine characteristics from several. It should also be noted that malingering is a situational, rather than static, response style (Rogers, 2018a; Rogers, Vitacco, & Kurus, 2010). Thus, feigning is best understood as a response style for specific circumstances instead of as a stable characteristic of the individual. Rogers (2018) wrote, “Some practitioners use- at least implicitly- the flawed logic ‘once a malingerer, always a malingerer.’ On the contrary, most efforts at malingering appear to be related to specific objectives in a particular context” (p. 8). Individuals may feign symptoms from one domain in one set of circumstances, and then may feign symptoms from a different domain or choose not to feign at all in a different set of circumstances.

Individuals receiving treatment or evaluation in forensic settings have numerous external incentives to feign or exaggerate symptoms of mental illness. Most significantly, this could include delaying adjudication, avoiding culpability, or reducing the sentence of a criminal offense. Other threats to validity in a forensic context include the importance of evaluations, the coerced nature of court referrals, and pressures from attorneys (Melton et al., 2018), as well as the possibility of transfer to a hospital or mental health unit, which is likely to be less secure and/or more comfortable than jail (Soliman & Resnick, 2010; Vitacco, 2018). In regards to the prevalence of feigning in forensic settings, clinician estimates range from 15% to 18%, and research findings range from less than 10% to over 25% (Boccaccini, Murrie, & Duncan, 2006). In a more recent review of the literature, Wygant and Lareau (2015) found that exaggeration of symptoms was estimated to occur 18 to 33% of the time in personal injury, medical, disability, and criminal cases. These estimates are troubling, particularly given the high stakes of forensic evaluations and the limited bed space available in state hospitals (Ray, 2009). Furthermore, although it has been posited that clients in therapy and treatment settings "have little incentive to practice conscious deception and manipulation" (Melton et al., 2018, p. 57), individuals in these non-forensic settings (i.e., clinical or counseling) may also be motivated by incentives such as obtainment of psychotropic medications, disability benefits, shelter, or avoidance of military service (Resnick & Knoll, 2018; Vitacco, 2018). Indeed, the results of surveys of over 500 forensic experts indicate that malingering is not rare in either clinical or forensic contexts (Rogers, 2018a). Melton and colleagues (2018) wrote:

> Given the significant potential for deception and the implications for the validity of their findings, mental health professionals should have a low threshold for suspecting less-than-candid responding. At the same time, given the limitations of science referenced above, and the weight that

> labels used to describe response styles (e.g., 'malingerer,' 'faker') carry with legal decisionmakers, the examiner should make sure that conclusions about an examinee's response style have a solid foundation. (p. 57)

Given that feigning is not a rare occurrence and is observed in multiple settings, it is important for clinicians to have both the knowledge and tools needed to accurately assess the veracity of reported symptoms. In order to make conclusions, diagnoses, and recommendations for treatment, evaluators first need to know whether the data collected is valid (Lezak et al., 2012). In forensic settings, mental health testimony is common in criminal proceedings involving culpability or sanity, competency to stand trial or to be executed, mitigation of sentencing, and intellectual disability (Haydt, 2015). In a survey of 102 forensic evaluators, respondents reported evaluating malingering or response style an average of 39.61 times in the previous year (McLaughlin & Kan, 2014). During evaluations of response style or malingering, the evaluators reported using forensic assessment instruments (FAIs) or forensically relevant instruments (FRIs) an average of 66.44% of the time, multiscale inventories 52.21% of the time, and cognitive or neuropsychological instruments 21.41% of the time.

Some experts have argued that forensic assessments should routinely include evaluation of response style or malingering. Rogers (2018) wrote, "When the outcome of an evaluation has important consequences, malingering should be systematically evaluated. Its professional neglect is a serious omission" (p. 8). However, understanding the importance of evaluating symptom veracity must be accompanied by evidence-based methods of doing so. A review of the malingering literature shows that clinicians and researchers use and recommend a variety of different methods in order to assess feigning, many of which are conflicting. The ability to detect feigned mental illness has been called an "advanced clinical skill," as the clinician must understand and be able to recognize the variety of phenomenologically disparate symptoms experienced by individuals with genuine psychological disorders (Resnick & Knoll, 2008).

Although some clinicians continue to believe in the superiority of clinical judgment over psychological assessments, a variety of factors have been demonstrated to limit the accuracy of unstructured clinical judgment (Borum, Otto, & Golding, 1993; Gottfried, Schenk, & Vitacco, 2016). Furthermore, no studies have been able to demonstrate that evaluators can detect malingering based solely on unstructured clinical interviews (Harris & Resnick, 2003; Melton et al., 2018) and the admissibility of expert testimony based solely on clinical interviews has been questioned (Melton et al., 2018). Unstructured clinical judgment is also subject to numerous memory and judgment biases (Melton et al., 2018) that can be reduced through the use of structured interviews or assessments. In comparison, numerous studies have demonstrated the utility of psychological assessments in the detection of feigned psychosis, affective symptoms, and cognitive impairment (Melton et al., 2018; Resnick & Knoll, 2018). These assessments include specialized response bias measures as well as broadband measures containing validity scales, indices, and discriminant functions. The most widely used and researched measures are summarized in the following section, followed by a more in-depth discussion of the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005).

Measures of Response Bias

Broadband measures. The Minnesota Multiphasic Personality Inventory- Second Edition (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kraemmer, 1989; Butcher et al., 2001) and the MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) are widely used and researched multi-scale measures of personality and psychopathology. The MMPI contains a number of validity and supplementary scales designed to assess the respondent's response style, including malingering, defensiveness, and inconsistent responding (Wygant, Walls, Brothers, & Berry, 2018). The scales most applicable to the detection of overreporting response styles include Infrequency/Infrequent Responses (F/F-r), Back Infrequency (F_B), Infrequency Psychopathology/Infrequent Psychopathology Responses (F_P/F_P-r), Fake Bad/Symptom Validity (FBS/FBS-r), Inconsistent Response (IR), Gough Dissimulation (Ds), Infrequent Somatic Responses (F_s), and Response Bias Scale (RBS). These validity scales, particularly F_p and F, have been found to differentiate between honest responding and feigning in both known-groups (Sellbom, Toomey, Wygant, Kucharski, & Duncan, 2010) and analogue simulation (Sellbom & Bagby, 2010) design studies.

The Personality Assessment Inventory (PAI; Morey, 1991, 2007) is a widely used multi-scale measure of personality and psychopathology. The PAI contains several validity and supplementary scales designed to assess response style. The validity scales pertaining most to the detection of overreporting response styles include Negative Impression Management (NIM), Malingering Index (MAL), Rogers Discriminant Function (RDF), Negative Distortion Scale (NDS), and Multiscale Feigning Index (MFI) (Boccaccini & Hart, 2018). In a thorough meta-analysis of studies investigating the ability of PAI scales to detect overreporting, NIM, MAL, and RDF were found to have large predictive effects (Hawes & Boccaccini, 2009).

The Millon Clinical Multiaxial Inventory- Fourth Edition (MCMI-IV; Millon, Grossman, & Millon, 2015) is another widely used multi-scale measure of personality and psychopathology. The MCMI-IV contains several validity scales pertaining to the detection of overreporting response styles, including the Validity Scale (V), Disclosure Scale (X), Desirability Scale (Y) and Debasement Scale (Z). Due to its recent publication, this latest revision of the MCMI has little research regarding its ability to detect feigned psychopathology. Furthermore, the literature on the previous edition (MCMI-III; Millon, Millon, Davis, & Grossman, 2006) was not promising, leading multiple researchers to question its ability to detect feigning or malingering (Berry & Schipper, 2007; Melton et al., 2018). A recent review of MCMI-IV validity scale research concluded that the shortcomings continue to apply and “practitioners should avoid using the MCMI-IV to differentiate between honest and dishonest responders in actual forensic cases” (Boccaccini & Hart, 2018, p. 293).

Specialized measures. The Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) and its second edition (SIRS-2; Rogers, Sewell, & Gillard, 2010) are widely used, well-validated measures of feigned mental disorders and are considered the gold standard for the classification of feigning by many malingering researchers (Rogers, 2018b). Melton and colleagues (2018) wrote, “The SIRS remains the most praised and best-validated measure” (p. 59) for evaluation of feigned psychopathology. The SIRS-2’s eight primary scales use validated detection strategies derived from the empirical literature. These scales include Rare Symptoms (RS), Symptom Combinations (SC), Improbable or Absurd Symptoms (IA), Blatant Symptoms (BL), Subtle Symptoms (SU), Selectivity of Symptoms (SEL), Severity of Symptoms

(SEV), and Reported vs. Observed (RO). Utilizing a flow chart of decision rules, the examinee's response style is classified into one of several categories: Genuine, Disengagement, Indeterminate-General, Indeterminate-Evaluate, or Feigning. The SIRS-2 manual lists sensitivity and specificity estimates of .80 and .975, respectively, for the classification of feigned mental disorders (Rogers, Sewell, & Gillard, 2010).

The Test of Memory Malingering (TOMM; Tombaugh, 1996) is a visual recognition measure designed to detect feigned memory impairment. The manual provides two cut scores; one based on below-chance performance and the other based on comparison to patients with genuine head injuries and cognitive impairments (Tombaugh, 1996). Validation studies have found the measure to be sensitive to feigning but insensitive to a wide array of genuine neurological impairments (Tombaugh, 1996, 2002).

The Validity Indicator Profile (VIP; Frederick, 1997) is validity measure designed to assess whether the examinee's performance on other cognitive measures should be considered a valid representation of his or her abilities. The VIP utilizes a two-alternative forced choice procedure over the course of 100 nonverbal and 78 verbal problems. Following administration, the items are reordered in terms of difficulty level, allowing the examinee's performance curve to be compared to the standard shape of the performance curve for compliant test-takers. Based on the performance curve measures and consistency measures, the examinee's performance is classified as either compliant (i.e., genuine) or one of three invalid response styles (Frederick & Crosby, 2000). In a review of the VIP's development and validation, effect sizes of test performance between individuals motivated to perform well and those motivated to perform poorly were large to very large (Frederick, 2002).

The Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001) is a brief structured interview, developed as a screening measure for feigned psychopathology in forensic settings (Smith, 2018). The 25 items on the M-FAST were written in an effort to utilize empirically-based detection strategies, such as those used in the SIRS (Miller, 2001). Its scales include Reported vs. Observed (RO), Extreme Symptomatology (ES), Rare Combinations (RC), Unusual Hallucinations (UH), Unusual Symptom Course (USC), Negative Image (NI), and Suggestibility (S). Research on the ability of the M-FAST to screen for feigning has generally been positive, with very large effect sizes and good negative predictive power (NPP) and sensitivity estimates (Smith, 2018). However, concerns have been noted regarding the over-reliance of the M-FAST on unlikely detection strategies with little use of amplified detection strategies (Rogers, Robinson, & Gillard, 2014).

The Structured Inventory of Malingered Symptomatology. The SIMS (Widows & Smith, 2005) was designed as a screening measure for feigning of both psychopathology and neuropsychology symptoms. In contrast to the M-FAST, the SIMS consists of 75 self-report, true-false items, and can be used in both clinical and forensic settings (Smith, 2018). The format of the SIMS scales differs from most other feigning measures in that the scales are organized into symptom types rather than detection strategies. In addition to a Total Score, this measure provides scores for Psychosis (P), Neurologic Impairment (NI), Amnestic Disorders (AM), Low Intelligence (LI), and Affective Disorders (AF) scales (Widows & Smith, 2005).

The SIMS manual (Widows & Smith, 2005) recommends a three-level interpretation of the measure, in which: (1) the Total Score is used for classification of suspected malingering, (2) scale scores are analyzed for qualitative data about response style, and (3) endorsement of individual items are used for additional information regarding the specific nature of the evaluee's reported symptoms. The recommended cut-scores for the SIMS Total Score and each SIMS scale are provided in Table 1. If an evaluee receives a score greater than the specified cut scores, he or she is identified as possibly feigning and should be referred for a full evaluation of response style.

Table 1

SIMS Scale Cut Scores as Recommended by the SIMS Manual

Scale	Cut Score
Psychosis (P)	> 1
Neurological Impairment (NI)	> 2
Amnestic Disorders (AM)	> 2
Low Intelligence (LI)	> 2
Affective Disorders (AF)	> 5
Total	> 14

In the original cross-validation study, the SIMS Total Score was found to be the best predictor of feigning, with sensitivity of .96 and specificity of .88. Sensitivity rates for the individual scales ranged from .75 (AF) to .88 (AM), while specificity rates ranged from .52 (LI) to .91 (AM; Widows & Smith, 2005). Subsequent known-groups research on the SIMS as a screening measure of feigning has been generally positive, revealing very large effect sizes and good NPP and sensitivity estimates. Among eight known-groups studies composed of forensic samples, sensitivity rates ranged from .71 to 1.00, NPP rates ranged from .75 to 1.00, and effect sizes ranged from $d = 1.10$ to 3.07 (Smith, 2018). Indeed, in his review of this measure, Smith wrote, "The SIMS is very effective at retaining potential malingerers for a full assessment of malingering" (p.458).

While the SIMS has a demonstrated strength of high sensitivity rates, some researchers have raised concerns regarding the potential for high rates of false positives (i.e., low specificity) on this measure among clients with genuine psychopathology. The primary concern with low specificity values is that honest-responding evaluees will be falsely classified into the possibly feigning range. For example, Edens, Otto, and Dwyer (1999) evaluated the SIMS' ability to detect feigned depression, psychosis, and cognitive impairment in a sample in which all participants responded honestly on one occasion and feigned symptoms on another occasion. Using the entire sample, these researchers found sensitivity rates for the SIMS Total Score and scale scores ranged from .68 (P) to .96 (Total) and specificity rates ranged from .71 (LI) to .92 (N). However, moderate correlations were found between SIMS scores and a measure of genuine symptomatology (i.e., the Symptom Checklist-90-Revised), and utility estimates of participants reporting current clinical distress resulted in decreased specificity rates [.52 (AF) to .85 (P)]. They concluded, "Our findings indicate that genuinely symptomatic persons are at risk for being identified as malingering on the SIMS" (p. 395), particularly using the AF scale (Edens, Otto, & Dwyer, 1999). Similarly, Rogers, Robinson, and Gillard (2014) investigated the SIMS in a sample of inpatients instructed to either respond honestly or to exaggerate their genuine symptoms. Using the manual-recommended cut-score of > 14, the SIMS Total Score obtained poor specificity (.28) and misclassified over two-thirds of the genuine responders as feigning (PPP = .70). Rogers and colleagues therefore recommended a significantly higher cut-score for the SIMS Total Score (> 44), which in their sample resulted in sensitivity of .60 and specificity of .98. In a study composed of four inmate samples, classification using the SIMS Total Score (> 14) resulted in sensitivity of .90 for simulators and .85 for suspected malingerers, and in sensitivity of .97 for controls but only .40 for genuine patients (Edens, Poythress, & Watkins-Clay, 2007). The researchers suggested that the recommended cut-scores were less accurate in the classification of feigning among mental health samples compared to general population samples. Finally, in a recent meta-analysis of the SIMS literature, van Impelen, Merckelbach, Jelicic, and Merten (2014) reviewed 41 studies using the SIMS to screen for feigning. They found effect sizes ranging from d = 1.1 to 3.0 in known-groups studies and from d = 0.5 to 4.7 in simulation studies. Mean sensitivity was .97 when using the standard Total Score cut-score of > 14, and .90 when using a more conservative cut-score of > 16. Mean specificity evidenced a great deal of variability between studies, ranging from .37 to .70 with both Total Score cut-scores. Van Impelen and colleagues concluded that the SIMS contains satisfactory sensitivity but substandard specificity, a common weakness found in screening measures designed to maximize detection.

Rogers, Robinson, and Gillard (2014) recently developed two additional SIMS scales using unlikely detection strategies. The Rare Symptoms (RS) scale includes 15 items referring to symptoms that are endorsed frequently by feigners but infrequently by genuine patients. The Symptom Combinations (SC) scale includes 13 pairs of items that are endorsed together frequently by feigners but infrequently by genuine patients. Table 2 includes cut-scores for these two new scales, in addition to the significantly higher Total Score cut-score, discussed above. In the validation of these scales, both RS and SC yielded very large effect sizes (d = 1.63 and d = 2.01, respectively). Using the same sample, the SIMS P, LI, NI, and total score yielded very large effect sizes (ranging from d = 1.35 to d = 1.71), but AF and AM produced smaller effect sizes (d =0.95 and d = 1.19, respectively). The authors noted that although the results are promising for the ability of the two new scales to detect feigning, additional validation is needed (Rogers, Robinson, & Gillard, 2014).

Table 2

SIMS Additional Scale Cut Scores

Scale	Cut Score
Rare Symptoms (RS)	>6
Symptom Combinations (SC)	>6
Total	>44

Consistency of Feigning Measure Scores Over Time

Given the significant need to differentiate patients genuinely experiencing symptoms from those feigning symptoms, and the consequences of both false negatives and false positives in this decision-making process, it is imperative that evaluators have access to well-validated feigning instruments. While a number of assessment tools have been developed specifically for this purpose, little is known about how repeated administrations of feigning instruments may affect scores or how common it is for scores to change over time. Feigning is not a static attribute, so an individual who feigns symptoms in one setting or during one time period may respond genuinely in another setting or at a different time (Rogers, 2018a; Rogers, Vitacco, & Kurus, 2010). Similarly, an individual may feign different types of symptoms (e.g., psychosis versus cognitive impairment) in different situations.

In addition to the possibility of the evaluee's response style changing over time, scores on feigning instruments may be impacted by changes in genuine symptoms of psychopathology or by measurement error due to repeated administrations. Thus, while research supports the ability of many feigning instruments to detect feigning at one time point, we have little information about whether the instruments remain effective when used again with the same patient. For example, Rogers and Bender (2003) made the following caution regarding repeated administrations of the SIRS:

> An important caution is that the SIRS has not been validated for repeat administrations, especially across brief intervals. We have observed several forensic cases in which in an expert, apparently dissatisfied with the results from an earlier expert, readministered the SIRS. One grave concern is whether the evaluatee had access to the results of the previous report (written or oral) or reasonably inferred this feedback from general comments made by his or her attorney. This type of specific feedback on past SIRS performance may invalidate subsequent administrations. (p. 118)

Rogers, Vitacco, and Kurus (2010) investigated the question of SIRS test-retest variability in a sample of 55 inpatients adjudicated not guilty by reason of insanity (NGRI). The authors noted that they focused on patients genuinely experiencing symptoms due to the expected variability in individuals feigning symptoms. After screening out participants with elevated M-FAST scores, the remaining participants were administered the SIRS on two intervals, approximately 10 days apart. The concordance rate for primary scales using unlikely detection strategies was 99.5%, and the concordance rate for scales using amplified detection strategies was 97.6%. The overall SIRS classification evidenced a 100% concordance rate, as all participants met criteria to be classified as “genuine” on both administrations. Rogers and colleagues concluded that the SIRS was highly stable across repeated administrations; however, they questioned whether other measures of feigned psychopathology would show similar stability.

This significance of this question is amplified when one considers the utility of being able to administer a brief screener rather than a full-length feigning assessment. Given the limited resources available in most forensic settings, it is often more practical to administer screening assessments. This allows the well validated but lengthy SIRS to be given to only those patients who elevate screening measures of feigning. The SIMS has been the subject of two studies investigating this question; however, both studies contained significant limitations. Cima and colleagues (2003) investigated the psychometric properties of the German version of the SIMS. In the pilot study described in the article, 18 undergraduate students each completed the measure on two occasions, six weeks apart. All participants were instructed to respond honestly. The researchers found a test-retest correlation of .97, indicating excellent stability over the two administrations. In a similar study, Merckelbach and Smith (2003) investigated the psychometric properties of the Dutch version of the SIMS. Twenty-four undergraduate students completed the measure on two occasions, three weeks apart. Again, all participants were instructed to respond honestly. These researchers obtained a lower, but still satisfactory, test-retest correlation of .72. While these studies provided some initial data about repeated administrations of the SIMS, both were limited by small sample sizes, the reliance on undergraduate participants, and the lack of a clinical control group.

In addition to the scarcity of research investigating repeated administrations of feigning screeners, there is not yet any research evaluating more than two administrations. Over the course of a patient or defendant's lifespan, he or she may require such assessments on multiple occasions. Without data indicating the effects of three, four, or even more administrations of the same screening measure, clinicians cannot truly know whether changes in scores should be attributed to changes in response style, genuine symptoms, or various types of measurement error. In the field of neuropsychology, serial testing (i.e., repeated administrations) is often used to track neuropsychological impairments and strengths over time and to investigate effectiveness of treatments (Heilbronner et al., 2010). Due to the frequency of this practice, the American Academy of Clinical Neuropsychology released an official position on serial neuropsychological assessments, noting a number of potential challenges raised by repeated administrations (Heilbronner et al., 2010). They wrote, in part:

> Neuropsychologists who conduct a re-assessment need to be mindful and knowledgeable about the variables potentially affecting change, including practice effects (i.e., gains related to prior exposure to a test), psychometric factors (test reliability), patient characteristics (demographics, state variables, fatigue, motivation) and disease state of the individual being assessed, and about resulting interpretive complications. (p. 1269)

Although feigning assessments have a different purpose than most neuropsychological assessments, repeated administrations of these two types of assessments raise many of the same concerns. Heilbronner and colleagues (2010) wrote, “There is an obvious need for more data on normal change trajectories for all types of measures with all types of demographic variables and patient groups” (p. 1274). Certainly, the same is true for feigning assessments.

The Current Study

Feigning of mental illness is a concern in a variety of clinical settings, but particularly in forensic settings, in which the veracity of reported symptoms can impact the outcomes of legal proceedings. While several assessments tools have been developed and studied specifically to detect feigned and exaggerated response styles, the research on multiple administrations of such assessments is extremely limited. For example, only two studies have examined these effects for the SIMS, and both are limited in generalizability to clinical populations due to small samples consisting entirely of undergraduate students. The purpose of this dissertation was to study patterns of scores of the same feigning screener administered multiple times over a 12-month period, thereby investigating the consistency of self-reported feigning across time. An additional area of investigation concerned whether changes in feigning scores were accompanied by corresponding changes in scores on measures of psychopathology or could be attributed to participant characteristics. Finally, comparisons were made between the manual-recommended cut score and alternative cut scores recommended in the literature, particularly between examinees with and without genuine mental illness.

This dissertation study involved secondary data analyses using the publicly available dataset from the study entitled, “Evaluation of the Psychological Effects of Administrative Segregation in Colorado, 2007-2010” (O’Keefe & Klebe, 2014). Participants in this study were administered numerous psychological assessments at baseline and again at 3, 6, 9, and 12 months after baseline. Their patterns of scores on the SIMS were evaluated over the five administrations in order to explore the patterns of self-reported symptoms over time.

Research questions investigated in this dissertation included:

1. How does using the SIMS manual recommended cut-score of >14 versus the Rogers and colleagues (2014) recommended cut-score of >44 impact which participants fall in the feigning range?
    a. What proportion of participants score in the feigning range at each time point, using each cut-score? Can participants be classified into clear subgroups based on their SIMS classifications over time, such as never feign, sometimes feign, or always feign?

   b. How does mental health needs classification impact the proportion of participants who score in the feigning range? Do inmates with mental illness tend to receive significantly higher SIMS scores compared to inmates without mental illness?

2. How much consistency is there in SIMS scores and classifications over time?
   a. What is the test-reliability of SIMS Total score and subscale scores among incarcerated offenders?
   b. To what extent do classifications based on an initial SIMS administration predict classifications on subsequent administrations?
   c. How do the SIMS subgroups (i.e., never, sometimes, or always feign) differ on measures of psychopathology (i.e., the PAS or BPRS)?
3. If SIMS classifications do not show consistency over time (i.e., if participants' scores vacillate between honest and feigning ranges), does this inconsistency correspond with:
   a. Changes in scores on measures of psychopathology (i.e., PAS or BPRS)?
   b. Other participant characteristics (i.e., administrative segregation, general population, or San Carlos Correctional facility; and mental health needs versus no mental health needs)?
   c. Regression to the mean (i.e., more extreme scores or classifications are closer to the mean upon reassessment)?

CHAPTER II

METHOD

Participants

Participants were 270 male inmates incarcerated in the Colorado Prison System who had a hearing for administrative segregation or a diversion placement between July 2007 and March 2009. Participants were from five subgroups within the prison system: inmates in administrative segregation with mental illness (n = 64), inmates in administrative segregation without mental illness (n = 63), inmates in general population with mental illness (n = 33), and inmates in general population without mental illness (n = 43). The fifth group consisted of inmates with mental illness placed in a mental health prison (San Carlos Correctional Facility) due to behavioral disruptions (n = 67). Mental illness classifications were based upon the existing classifications used by the prisons, which included diagnoses, symptoms, and resources used. Exclusionary criteria included scheduled release from prison prior to completion of the study, low reading ability, or an inability to read English. Participants were also selected based on timing or location to other possible participants.

The initial screening group included 1,087 inmates; of these, 526 were excluded based on the exclusionary criteria. Of the 561 eligible inmates, 302 were approached for study participation and 270 inmates consented to participate and completed the baseline (Time 1) assessments. Of the original 270 participants, 261 completed Time 2 (3 months), 253 completed Time 3 (6 months), 243 completed Time 4 (9 months), and 236 completed Time 5 (12 months). In addition, 106 participants from the administrative segregation subgroups completed a sixth testing session due to delays caused by the waitlist for administrative segregation beds, during which they were held in a punitive segregation bed at their original facility. In both the original researchers' study and the current study, testing completed at each time point was kept in chronological order, such that the extra time point was added to the end (i.e., Time 6) rather than the beginning. Data collected during Time 6 was not used in the current study due to difficulties comparing measures completed at different times. See Table 3 for attrition rates between groups at each time point.

Table 3

Attrition at Each Time Point

Group	Time 1 Baseline	Time 2 3 Months	Time 3 6 Months	Time 4 9 Months	Time 5 12 Months	Time 6 (AS Only)
AS + MI	64	62	60	60	56	51
AS – MI	63	60	58	56	56	54
GP + MI	33	33	32	29	29	0
GP – MI	43	41	41	39	38	0
SCCF	67	65	62	59	57	0
Total	270	261	253	243	236	105

Note. AS = Administrative segregation housing, GP = General population housing, + MI = with mental illness, - MI = no mental illness, SCCF = San Carlos Correctional Facility

Some participants agreed to participate, then declined to participate at one or more time points, but then agreed to participate at later time points. Across the five time points, eight participants (3.0%) completed the SIMS one time, eight participants (3.0%) completed the SIMS two times, 11 participants (4.1%) completed the SIMS three times, 11 participants (4.1%) completed the SIMS four times, and 232 participants (85.9%) completed the SIMS all five times.

The full sample of 270 participants ranged in age from 17 to 59 ($M = 31.8$, $SD = 9.1$). Regarding racial/ethnic backgrounds, 40% of the sample was Caucasian, 36% Hispanic/Latino, 18% African-American, 4% Native American, and 1% Asian. Among the participants who were classified as having mental health needs, 56% were diagnosed with a serious and pervasive mental illness.

Procedures

The data for the study were initially collected as part of a larger study entitled, "Evaluation of the Psychological Effects of Administrative Segregation in Colorado, 2007-2010" (O'Keefe & Klebe, 2014). Prior to participation, the field researcher met individually with each inmate to explain the purpose, voluntary nature, risks, benefits, and compensation for the study, and to obtain informed consent. Participants received $10.00 compensation for participation in each testing session, for a maximum of $60.00 over the course of the study.

At the time of consent, the participants were administered a battery of twelve self-report psychological and cognitive assessments, as well as two prison staff-rated assessments of behavior and psychological functioning. Most of the assessments were re-administered every three months after baseline, resulting in follow-up testing sessions at 3, 6, 9, and 12 months. As detailed above, some participants completed a sixth testing session; however, this data were not used in the current study. Measures assessing traits of personality disorders, self-harm behaviors, and history of traumatic events were not re-administered due to the relative stability of these constructs.

Measures

Structured Inventory of Malingered Symptomatology (SIMS). The SIMS (Widows & Smith, 2005) is a screening measure of feigned psychopathological and neuropsychological symptoms. The measure consists of 75 true-false items and provides a Total Score as well as scores for Psychosis (P), Neurologic Impairment (NI), Amnestic

Disorders (AM), Low Intelligence (LI), and Affective Disorders (AF) scales. Two additional scales, Rare Symptoms (RS) and Symptom Combinations (SC), were developed by Rogers, Robinson, and Gillard (2014) in an effort to use the SIMS to screen for feigning using unlikely detection strategies.

The original researchers of this study provided item-level SIMS data for at all five time points. The current study calculated and used the SIMS Total Score, five scale scores, and two additional scale scores at each time point. The SIMS manual (Widows & Smith, 2005) provides cut-scores to suggest malingering for each of these scales. Consistent with the manual, this study used cut-scores of > 14 for the Total Score, > 1 for the P scale, > 2 for the NI, AM, and LI scales, and > 5 for the AF scale. For the two additional scales, this study used cut-scores of > 6 for both RS and SC, as recommended by Rogers and colleagues (2014). Finally, this study investigated the higher Total Score cut-score of > 44 recommended by Rogers and colleagues.

Brief Psychiatric Rating Scale- Expanded Version (BPRS-E). The BPRS-E (Ventura, Lukoff, Nuechterlein, Liberman, Green, & Shaner, 1993) is an expanded version of the original BPRS (Overall & Gorham, 1962; Overall & Klett, 1972), a measure designed to rapidly assess changes in psychiatric symptoms among inpatient populations. The BPRS-E consists of 24 items organized into a semi-structured interview, assessing a variety of affective and psychotic symptoms. Eleven items are clinician-rated based on the individual's responses to interview (i.e., somatic concern, anxiety, depression, suicidality, guilt, hostility, grandiosity, suspiciousness, hallucinations, unusual thought content, and disorientation), three items are based on both the patient's responses and clinician observations during the interview (i.e., elevated mood, bizarre behavior, and self-neglect), and ten items are based entirely on clinician observations (conceptual disorganizations, blunted affect, emotional withdrawal, motor retardation, tension, uncooperativeness, excitement, distractibility, motor hyperactivity, and mannerisms and posturing). Each item is rated on a scale from 1 "not present" to 7 "extremely severe," resulting in Total Scores ranging from 24 to 168.

The factor analytic structure of the BPRS has been widely studied, with most studies reporting a four or five factor solution in clinical patient populations. Burlingame and colleagues (2006) reported the most common factor components of the BPRS-E to be thought disturbance/positive symptoms, anxiety/depression, withdrawal/negative symptoms, hostile suspicious/paranoid, and activity/mania. A recent meta-analysis of BPRS-E factor analytic studies found four components to be most common: affect (i.e., depression, anxiety, guilt, suicidality), positive symptoms (i.e., hallucinations, unusual thought content, grandiosity, suspiciousness), negative symptoms (i.e., blunted affect, withdrawal, motor retardation), and activation (i.e., excitement, elevated mood, hyperactivity, distractibility; Dazzi, Shafer, & Lauriola, 2016). The researchers also noted that a fifth factor of disorganization (i.e., disorganization, disorientation, mannerisms, self-neglect) was not statistically supported but was present in some analyses studied.

The original researchers provided item-level BPRS data and Total Scores at Time 1, Time 3, and Time 5. In addition, the dataset included scores at each of these same time points for a Withdrawal composite, Anxious-Depressed composite, Activation/Activity composite, Thought Disorders composite, and Hostility-Suspiciousness composite. The current study used the BRPS Total Score at each time point available, as well as the Thought Disorders and Anxious-Depressed composite scores. These two composites were chosen due to their relevance in symptoms to the Psychosis and Affective Disorders subscales on the SIMS.

Personality Assessment Screener (PAS). The PAS (Morey, 1997) is a self-report screening version of the Personality Assessment Inventory (PAI; Morey, 1991, 2007), designed to evaluate the need for a full evaluation and to target particular areas of clinical interest. The 22 PAS items consist of the 22 PAI items found to be most sensitive to clinical problems. These items are organized into a Total Score and 10 clinical problem domains, including Negative Affect, Acting Out, Health Problems, Psychotic Features, Social Withdrawal, Hostile Control, Suicidal Thinking, Alienation, Alcohol Problem, and Anger Control. Items are rated on a 4-point scale from 0 “false, not at all true” to 3 “very true,” resulting in Total Scores ranging from 0 to 66. In addition to raw scores, the measure provides probability (p) values, which indicate the probability that the examinee would obtain a clinically significant score (i.e., $T \geq 70$) on the full PAI.

Research examining the utility of the PAS in screening for emotional and behavioral dysfunction has generally been positive, although the recommended cut-scores for interpretation are still debated. In one study, researchers investigated the predictive accuracy of the PAS Total Score in the identification of PAI profiles with at least one clinically significant scale elevation among of male and female participants either incarcerated or receiving court-mandated substance abuse treatment (Kelley, Edens, & Douglas, 2018). They obtained AUC values of .88 (SE = .01) using community norms and .83 (SE = .01) using correctional norms. Kelley and colleagues also suggested using a PAS Total score cut-score of ≥ 29 (rather than the manual-recommended cut-score of ≥ 19) to increase specificity, as 92% of the sample scored greater than 19 using community norms and 51% scored greater using correctional norms (Kelley, Edens, & Douglas, 2018). Another study of the PAS' predictive accuracy in identifying PAI elevations was composed of three male correctional samples (Edens, Penson, Smith, & Ruchensky, 2018). Similar to the prior study, the combined samples obtained *AUC* values of .85 (*SE* = .02) using community norms and .81 (*SE* = .02) using correctional norms. The manual recommended cut-score of ≥ 19 resulted in sensitivity of .89 and specificity of .37, while a more conservative cut-score of ≥ 29 resulted in lower sensitivity (.60) but higher specificity (.92; Edens et al., 2018). The PAS has also evidenced utility in screening for psychological dysfunction in a variety of non-offender samples, such as with veterans (Creech et al., 2010), in urban primary care settings (Porcerelli et al, 2012), and with individuals exposed to childhood abuse and/or intimate partner violence (Porcerelli et al., 2015).

The original researchers provided item-level PAS data at all five time points. The current study calculated and used the PAS Total Score at each time point available, as well as the Psychotic Features and Negative Affect scores. These two clinical problem domains were chosen due to their relevance in symptoms to the Psychosis and Affective Disorders subscales on the SIMS.

CHAPTER III

RESULTS

Research Question 1a

For my investigation of the two recommended SIMS cut-scores (i.e., > 14 versus > 44), I began by calculating the number of inmates who scored in the feigning range for each of the five time points using each cut-score. As can be seen in Table 4, the cut-score of > 14 resulted in 35.2% to 44.4% of examinees scoring in the feigning range, while the cut-score of > 44 resulted in 0.7% to 3.1% of examinees scoring in the feigning range at each time point.

Table 4

Number of Participants in Feigning Range at Each Time Point

Cut Score	Time 1 *N* = 270	Time 2 *N* = 261	Time 3 *N* = 250	Time 4 *N* = 243	Time 5 *N* = 236
SIMS Total > 14	120 (44.4%)	105 (40.2%)	103 (41.2%)	100 (41.2%)	83 (35.2%)
SIMS Total > 44	2 (0.7%)	8 (3.1%)	5 (2.0%)	7 (2.9%)	6 (2.5%)

For each recommended SIMS cut score, I sorted participants into one of three feigning subgroups (Never Feigned, Sometimes Feigned, Always Feigned) based on the proportion of times they scored in the feigning range across administrations (see Table 5). Using a cut score of > 14, 41.5% were placed in the Never Feigning category, 36.3% were placed in the Sometimes Feigning category, and 22.2% were placed in the Always Feigning category. Thus, using a cut score of > 14, 58.5% of the participants were classified as feigning on at least one occasion. Using a cut score of > 44, 95.6% were placed in the Never Feigning category, 4.1% were placed in the Sometimes Feigning category, and 0.4% were placed in the Always Feigning category. Thus, using a cut score of > 44, only 4.5% were classified as feigning on at least one occasion.

Table 5

Feigning Groups Based on Proportion of Times in Feigning Range (N = 270)

Cut Score	Never	Sometimes	Always
SIMS Total > 14	112 (41.5%)	98 (36.3%)	60 (22.2%)
SIMS Total > 44	258 (95.6%)	11 (4.1%)	1 (0.4%)

Research Question 1b

I next investigated the two recommended cut-scores separately for individuals with and without a documented mental illness, all without any known reason to feign when completing the study's measures (see Tables 6-9). Using a cut-score of > 14, 75.5% of inmates with mental illness scored in the feigning range at least once, and 34.2% scored in the feigning range across all five administrations. For inmates without mental illness, 35.7% scored in the feigning range at least once, and 6.1% scored in the feigning range across all administrations. In other words, the odds of a person with mental illness scoring in the feigning range at least once were 5.56 times the odds of a participant without mental illness scoring in the feigning range at least once, a large (OR = 5.56, 95% CI [3.28, 9.43]) and statistically significant difference, $\chi2$ (N = 270) = 43.15, p < .001.

Using a cut-score of > 44, only 6.4% of inmates with mental illness scored in the feigning range on at least one occasion, and only 0.6% scored in the feigning range across all five administrations. For inmates without mental illness, 1.7% scored in the feigning range at least once, and 0.0% scored in the feigning range across all administrations. Tables 6 through 9 provide the specific proportions and numbers of participants for inmates with and without mental illness. With this cut score, the odds of a person with mental illness scoring in the feigning range at least once were 3.90 times the odds of a participant without mental illness scoring in the feigning range at least once, a medium (OR = 3.90, 95% CI [0.84, 18.14]) but non-significant difference, $\chi2$ (N = 270) = 3.45, p = .06.

Table 6

MI Group: Number of Participants in Feigning Range at Each Time Point

Cut Score	Time 1 N = 155	Time 2 N = 151	Time 3 N = 144	Time 4 N = 139	Time 5 N = 134
SIMS Total > 14	94 (60.6%)	86 (57.0%)	84 (58.3%)	79 (56.8%)	66 (49.3%)
SIMS Total > 44	2 (1.3%)	7 (4.6%)	5 (3.5%)	6 (4.3%)	6 (4.5%)

Table 7

MI Group: Proportion of Times in Feigning Range (n = 155)

Cut Score	Never	Sometimes	Always
SIMS Total > 14	38 (24.5%)	64 (41.3%)	53 (34.2%)
SIMS Total > 44	145 (93.5%)	9 (5.8%)	1 (0.6%)

Table 8

No MI Group: Number of Participants in Feigning Range at Each Time Point

Cut Score	Time 1 N = 115	Time 2 N = 110	Time 3 N = 106	Time 4 N = 104	Time 5 N = 102
SIMS Total > 14	26 (22.6%)	19 (17.3%)	19 (17.9%)	21 (20.2%)	17 (16.7%)
SIMS Total > 44	0 (0.0%)	1 (0.9%)	0 (0.0%)	1 (1.0%)	0 (0.0%)

Table 9

No MI Group: Proportion of Times in Feigning Range (n = 115)

Cut Score	Never	Sometimes	Always
SIMS Total > 14	74 (64.3%)	34 (29.6%)	7 (6.1%)
SIMS Total > 44	113 (98.3%)	2 (1.7%)	0 (0.0%)

I used independent samples t-tests to compare mean SIMS Total Scores for inmates with and without mental illness. At all time points, inmates with mental illness received significantly higher mean Total Scores compared to inmates without mental illness (t (268) = 8.82, d = 1.08; t (259) = 7.09, d = 0.89; t (248) = 7.98, d = 1.02; t (241) = 6.52, d = 0.84; and t (234) = 6.35, d = 0.83, respectively; p < .001 for all). Mean scores for the no mental illness group ranged from 9.49 to 10.23 (SD = 5.53 to 7.46); mean scores for the mental illness group ranged from 17.79 to 19.94 (*SD* = 10.75 to 12.20). Table 10 contains means and standard deviations for each of these groups, across the five time points.

Table 10

Comparing Mean SIMS Total Scores for Participants with and without Mental Illness

Group	Time 1	Time 2	Time 3	Time 4	Time 5
Mental Illness	19.94	19.22	19.59	18.90	17.79
(n = 155)	(10.75)	(12.05)	(11.91)	(12.20)	(12.17)
No Mental Illness	10.23	10.00	9.65	10.18	9.49
(n = 115)	(5.68)	(7.46)	(5.53)	(7.04)	(5.84)

Note. Numbers indicate *M* (*SD*)

Given the discrepancy in SIMS scores found between participants with and without mental health needs, I calculated correlations between SIMS scores and PAS/BPRS scores at each time point. As can be seen in Table 11, SIMS Total Scores significantly correlated with PAS Total Scores and BPRS Total Scores at all time points, with generally moderate-sized correlations.

Table 11

Correlations between SIMS Total Scores and Scores on Measures of Psychopathology

Time Point	PAS Total Score	BPRS Total Score
Baseline/Time 1	.50**	.40**
Time 2	.43**	--
Time 3	.55**	.47**
Time 4	.47**	--
Time 5	.45**	.31**

Note: * $p < .05$, ** $p < .01$

Research Question 2a

Next, I investigated test-retest reliability of SIMS scores for the sample as a whole, examining Pearson correlations between SIMS Total Scores at all combinations of time points. As seen in Table 12, correlations ranged from r = .61 (Time 1 - Time 5) to r = .83 (Time 3 - Time 5), suggesting moderate to high reliability. I also calculated Pearson correlations between scores for each SIMS scale at all combinations of time points (see Tables 13 - 19). Correlations for NI ranged from r = .57 to .81, AF from r = .49 to .68, P from r = .52 to .82, LI from r = .60 to .74, AM from r = .55 to .75, RS from r = .55 to .76, and SC from r = .41 to .57. Thus, most SIMS scales were comparable in reliability, with relative weaknesses in reliability for the affective disorders and symptom combinations scales.

Table 12

Test-Retest Reliability of the SIMS Total Score: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	15.80	10.15	1.00				
Time 2	261	15.33	11.31	.70**	1.00			
Time 3	250	15.38	10.89	.71**	.79**	1.00		
Time 4	243	15.17	11.17	.63**	.72**	.82**	1.00	
Time 5	236	14.20	10.74	.61**	.68**	.83**	.79**	1.00

Note: * $p < .05$, ** $p < .01$.

Table 13

Test-Retest Reliability of the SIMS NI Scale: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	2.96	2.71	1.00				
Time 2	261	2.81	3.00	.64**	1.00			
Time 3	250	2.79	2.85	.57**	.66**	1.00		
Time 4	243	2.78	2.94	.59**	.64**	.78**	1.00	
Time 5	236	2.58	2.95	.59**	.59**	.81**	.75**	1.00

Note: * $p < .05$, ** $p < .01$.

Table 14

Test-Retest Reliability of the SIMS AF Scale: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	5.47	2.56	1.00				
Time 2	261	5.35	2.78	.57**	1.00			
Time 3	251	5.62	2.70	.55**	.65**	1.00		
Time 4	243	5.45	2.64	.49**	.62**	.66**	1.00	

Time Point	N	M	SD	Time 1	Time 2	Time 3	Time 4	Time 5
Time 5	236	5.28	2.50	.53**	.58**	.68**	.65**	1.00

Note: * $p < .05$, ** $p < .01$.

Table 15

Test-Retest Reliability of the SIMS P Scale: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	2.36	3.09	1.00				
Time 2	261	2.10	2.93	.66**	1.00			
Time 3	250	2.08	2.84	.67**	.76**	1.00		
Time 4	243	2.10	3.03	.59**	.66**	.82**	1.00	
Time 5	236	1.82	2.86	.52**	.61**	.77**	.76**	1.00

Note: * $p < .05$, ** $p < .01$.

Table 16

Test-Retest Reliability of the SIMS LI Scale: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	2.44	1.89	1.00				
Time 2	261	2.37	1.93	.61**	1.00			
Time 3	251	2.43	2.11	.62**	.66**	1.00		
Time 4	243	2.32	2.09	.60**	.64**	.71**	1.00	
Time 5	236	2.34	2.04	.61**	.67**	.66**	.74**	1.00

Note: * $p < .05$, ** $p < .01$.

Table 17

Test-Retest Reliability of the SIMS AM Scale: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	2.57	2.90	1.00				
Time 2	261	2.70	3.28	.62**	1.00			
Time 3	251	2.45	3.14	.67**	.75**	1.00		
Time 4	243	2.51	3.34	.55**	.69**	.76**	1.00	
Time 5	236	2.19	3.07	.57**	.66**	.77**	.74**	1.00

Note: * $p < .05$, ** $p < .01$.

Table 18

Test-Retest Reliability of the SIMS RS Scale: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	1.63	1.90	1.00				
Time 2	261	1.46	1.91	.63**	1.00			

Time 3	249	1.56	2.07	.67**	.76**	1.00		
Time 4	243	1.48	2.04	.55**	.65**	.74**	1.00	
Time 5	236	1.48	1.82	.61**	.65**	.70**	.69**	1.00

Note: * $p < .05$, ** $p < .01$.

Table 19

Test-Retest Reliability of the SIMS SC Scale: Descriptive Statistics and Correlations

				Correlations				
Time Point	*N*	*M*	*SD*	Time 1	Time 2	Time 3	Time 4	Time 5
Baseline/Time 1	270	3.92	2.44	1.00				
Time 2	261	3.44	2.18	.45**	1.00			
Time 3	246	3.33	2.21	.52**	.50**	1.00		
Time 4	243	3.18	2.09	.41**	.57**	.55**	1.00	
Time 5	236	3.15	2.20	.43**	.47**	.48**	.56**	1.00

Note: * $p < .05$, ** $p < .01$.

Research Question 2b

I used classification accuracy statistics (i.e., sensitivity, specificity, positive predictive power (PPP), negative predictive power (NPP), and overall accuracy) to examine the consistency of SIMS classifications across time. Two sets of analyses were included, the first focusing on changes between the baseline administration and subsequent administrations, and the second focusing on changes between consecutive administrations. It should be noted that in this paper, classification accuracy statistics were used somewhat differently than in most malingering research. Rather than using a test to predict a known group (i.e., using the SIMS feigning classification to predict true malingering/honest status), these statistics used the SIMS classification at Time 1 to predict the SIMS classification at a later time point. Using the Time 1-Time 2 comparison as an example, sensitivity in this context indicates the probability that participants feigned at Time 1 when they feigned at Time 2 (i.e., true positive rate), specificity indicates the probability that participants did not feign at Time 1 when they did not feign at Time 2 (i.e., true negative rate), PPP indicates the probability that participants feigned at Time 2 when they feigned at Time 1, NPP indicates the probability that participants did not feign at Time 2 when they did not feign at Time 1, and accuracy indicates the overall probability that participants were correctly classified at Time 1 using Time 2 as the criterion.

For the first set of analyses, classification accuracy statistics were calculated for the ability of the baseline SIMS classification to predict classifications on subsequent administrations (i.e., Time 1 vs. Time 2, Time 1 vs. Time 3, Time 1 vs. Time 4, Time 1 vs. Time 5). Table 20 contains the cell counts for each of these classification comparisons, while Table 21 contains classification accuracy statistics. Sensitivity rates ranged from 77.11% to 81.55%, specificity rates from 73.86% to 79.59%, PPP from 61.54% to 73.86%, NPP from 84.33% to 86.03%, and overall accuracy rates from 75.00% to 80.40%. McNemar's test indicated the proportion of participants who scored in the feigning range at Time 1 and the proportion who scored in the feigning range at Time 2 were not significantly different ($p = .10$, $OR = 1.65$, 95% CI [0.95, 2.88]). Additional McNemar's tests revealed nonsignificant differences between Time 1 and Time 3 ($p = .15$, $OR = 1.58$, 95% CI [0.89, 2.81]), and Time 1 and Time 4 ($p = .26$, $OR = 1.50$, 95% CI [0.82, 2.50]); however, the proportions of participants who scored in the feigning range at Time 1 and Time 5 were significantly different ($p < .01$, $OR = 2.11$, 95% CI [1.22, 3.63]). This suggests that the proportions of participants who score in the feigning range at two different time points are more likely to be significantly different when the time points are farther apart.

Table 20

Cell Counts for Ability of Time 1 Classification to Predict Subsequent Classifications

Time 1	Time 2		Time 3		Time 4		Time 5	
	Feign	Honest	Feign	Honest	Feign	Honest	Feign	Honest
Feign	85	33	84	30	79	30	64	40
Honest	20	123	19	117	21	113	19	113

Table 21

Classification Accuracy Statistics for Subsequent Classifications

Time Point Comparison	Sensitivity (CI)	Specificity (CI)	PPP (CI)	NPP (CI)	Accuracy (CI)
Time 1-Time 2	80.95% (72.13-87.96)	78.85% (71.59-84.97)	72.03% (65.23-77.95)	86.01% (80.44-90.19)	79.69% (74.30-84.40)
Time 1-Time 3	81.55% (72.70-88.51)	79.59% (72.17-85.79)	73.68% (66.76-79.61)	86.03% (80.27-90.31)	80.40% (74.93-85.13)
Time 1-Time 4	79.00% (69.71-86.51)	79.02% (71.43-85.38)	72.48% (65.35-78.62)	84.33% (78.47-88.82)	79.01% (73.35-83.96)
Time 1-Time 5	77.11% (66.58-85.62)	73.86% (66.15-80.62)	61.54% (54.46-68.16)	85.61% (79.85-89.92)	75.00% (68.97-80.39)

Note. CI = 95% confidence interval.

For the second set of analyses, classification accuracy statistics were calculated for the ability of each SIMS administration to predict classifications on each consecutive administration (i.e., Time 1 vs. Time 2, Time 2 vs. Time 3, Time 3 vs. Time 4, Time 4 vs. Time 5). Table 22 contains the cell counts for each of these classification comparisons, while Table 23 contains classification accuracy statistics. Sensitivity rates ranged from 80.39% to 84.34%, specificity rates from 78.85% to 87.07%, PPP from 72.03% to 81.19%, NPP from 86.01% to 90.78%, and accuracy rates from 79.69% to 84.34%. McNemar's tests revealed nonsignificant differences between the proportions of participants who scored in the feigning range at Time 1 and Time 2 ($p = .10$, $OR = 1.65$, 95% CI [0.95, 2.88]), Time 2 and Time 3 ($p = 1.00$, $OR = 1.05$, 95% CI [0.56, 1.97]), Time 3 and Time 4 (p = .87, OR = 1.11, 95% CI [0.59, 2.20]), and Time 4 and Time 5 (p = .73, OR = 1.92, 95% CI [0.98, 3.76]). These results suggest that similar proportions of participants are likely to score in the feigning range at consecutive time points.

Table 22

Cell Counts for Ability of Each Classification to Predict the Consecutive Classification

	Time 2		Time 3		Time 4		Time 5	
	Feign	Honest	Feign	Honest	Feign	Honest	Feign	Honest
Time 1								
Feign	85	33	-	-	-	-	-	-
Honest	20	123	-	-	-	-	-	-
Time 2								
Feign	-	-	82	19	-	-	-	-
Honest	-	-	20	128	-	-	-	-
Time 3								
Feign	-	-	-	-	80	20	-	-
Honest	-	-	-	-	18	121	-	-
Time 4								
Feign	-	-	-	-	-	-	70	25
Honest	-	-	-	-	-	-	13	128

Table 23

Classification Accuracy Statistics for Consecutive Classifications

Time Point Comparison	Sensitivity (CI)	Specificity (CI)	PPP (CI)	NPP (CI)	Accuracy (CI)
Time 1- Time 2	80.95% (72.13-87.96)	78.85% (71.59-84.97)	72.03% (65.23-77.95)	86.01% (80.44-90.19)	79.69% (74.30-84.40)
Time 2- Time 3	80.39% (71.35-87.59)	87.07% (80.55-92.04)	81.19% (73.73-86.91)	86.49% (81.13-90.50)	84.34% (79.22-88.62)
Time 3- Time 4	81.63% (72.53-88.74)	85.82% (78.95-91.12)	80.00% (72.50-85.85)	87.05% (81.50-91.12)	84.10% (78.84-88.50)
Time 4- Time 5	84.34% (74.71-91.39)	83.66% (76.83-89.14)	73.68% (65.91-80.22)	90.78% (85.61-94.22)	83.90% (78.58-88.35)

Note. CI = 95% confidence interval.

Research Question 2c

I next used six mixed-model repeated measures MANOVAs to compare the three subgroups of inmates (i.e., Never, Sometimes, or Always Feigning using the >14 SIMS cut-score) on measures of psychopathology (i.e., PAS and BPRS) across the five time points. For all analyses, feigning group served as the between-subjects independent variable and time point served as the within-subjects independent variable. Dependent variables were the various PAS and BRPS scale scores at each time point. PAS analyses included all five time points, while BRPS analyses included the three time points available (i.e., Time 1, Time 3, and Time 5). Table 24 contains descriptive statistics for all MANOVAs.

Table 24

Descriptive Statistics for Measures of Psychopathology by Feigning Group

	Time 1 M (SD)	Time 2 M (SD)	Time 3 M (SD)	Time 4 M (SD)	Time 5 M (SD)
PAS Total Score					
Never (*n* = 84)	25.51 (7.44)	24.64 (8.12)	24.23 (7.96)	23.83 (7.53)	24.33 (7.88)
Sometimes (*n* = 81)	32.36 (8.97)	31.02 (8.51)	30.23 (7.40)	30.14 (8.86)	29.63 (7.85)
Always (*n* = 43)	35.49 (9.46)	35.40 (9.17)	36.12 (8.02)	35.00 (8.26)	34.23 (9.47)
Total (*n* = 208)	30.24 (9.38)	29.35 (9.44)	29.02 (8.95)	28.60 (9.25)	28.44 (9.01)
PAS Psychotic Features					
Never (*n* = 92)	1.21 (1.36)	1.00 (1.21)	1.11 (1.31)	1.09 (1.18)	1.00 (1.13)
Sometimes (*n* = 88)	1.98 (1.71)	1.64 (1.58)	1.76 (1.72)	1.69 (1.71)	1.82 (1.84)
Always (*n* = 48)	2.25 (1.79)	2.46 (1.98)	2.63 (1.62)	2.31 (1.80)	2.54 (1.77)
Total (*n* = 228)	1.55 (1.63)	1.55 (1.63)	1.68 (1.64)	1.58 (1.60)	1.64 (1.68)
PAS Negative Affect					
Never (*n* = 90)	3.24 (1.97)	3.23 (2.02)	2.99 (1.83)	2.97 (1.88)	2.88 (1.77)
Sometimes (*n* = 86)	4.90 (2.30)	4.10 (2.03)	3.91 (1.99)	4.17 (2.22)	4.01 (2.10)
Always (*n* = 46)	5.09 (2.19)	5.33 (2.26)	5.54 (2.11)	5.50 (2.24)	5.22 (2.33)
Total (*n* = 222)	4.27 (2.30)	4.00 (2.21)	3.87 (2.16)	3.96 (2.29)	3.80 (2.20)
BRPS Total Score					
Never (*n* = 71)	30.41 (6.47)	-	28.72 (6.49)	-	28.70 (5.67)
Sometimes (*n* = 76)	36.11 (9.30)	-	31.68 (5.25)	-	31.49 (6.02)
Always (*n* = 42)	37.33 (8.10)	-	36.71 (7.39)	-	34.64 (6.49)
Total (*n* = 189)	34.24 (8.57)	-	31.69 (6.90)	-	31.14 (6.37)
BPRS Anxious-Depres.					
Never (*n* = 71)	7.72 (2.34)	-	7.15 (2.57)	-	7.13 (2.60)

Sometimes (n = 76)	9.34 (3.31)	-	8.14 (2.55)	-	8.08 (2.48)
Always (n = 41)	10.37 (3.36)	-	9.78 (2.41)	-	9.73 (3.52)
Total (n = 188)	8.95 (3.15)	-	8.13 (2.70)	-	8.08 (2.93)
BPRS Thought Disord.					
Never (n = 71)	5.54 (1.29)	-	5.35 (1.08)	-	5.41 (1.32)
Sometimes (n = 76)	7.01 (3.11)	-	5.79 (1.34)	-	6.11 (2.16)
Always (n = 41)	7.61 (2.87)	-	7.12 (2.04)	-	6.51 (1.91)
Total (n = 188)	6.59 (2.64)	-	5.91 (1.58)	-	5.93 (1.87)

Note. PAS = Personality Assessment Screener, BRPS = Brief Psychiatric Rating Scale. Some cases were not included due to listwise deletion of missing data. BPRS data was only available at time points 1, 3, and 5.

For the first MANOVA, the PAS Total Scores at Times 1 through 5 served as the dependent variables. There were significant main effects for time point (F (4, 202) = 2.57, p = .04, Wilk's Λ = 0.95, partial η^2 = .05) and feigning group (F (2, 205) = 35.24, p < .001, partial η^2 = .26). The time point/feigning group interaction effect was not significant (F (8, 404) = 0.85, p = .56, Wilk's Λ = 0.97, partial η^2 = .02). For the time point main effect, pairwise comparisons indicated that the only significant difference was between Time 1 and Time 5 (p = .02). Participants received higher scores at Time 1 (M = 30.24, SD = 9.38) compared to Time 5 (M = 28.44, SD = 9.01). For the feigning group main effect, a Tukey post hoc test revealed that the mean PAS Total Scores in the Never group were significantly lower than both the Sometimes (p < .001) and Always (p < .001) groups, and the Sometimes group scored significantly lower than the Always group (p < .01).

For the second MANOVA, the PAS Psychotic Features subscale scores at Times 1 through 5 served as the dependent variables. There was a significant main effect for feigning group (F (2, 225) = 20.71, p < .001, partial η^2 = .16), but the main effect for time point was not significant (F (4, 222) = 0.95, p = .44, Wilk's Λ = 0.98, partial η^2 = .02). The interaction effect was also not significant (F (8, 444) = 0.85, p = .56, Wilk's Λ = 0.97, partial η^2 = .02). For the feigning group main effect, a Tukey post hoc test revealed that the mean PAS Total Scores in the Never group were significantly lower than both the Sometimes ($p < .001$) and Always ($p < .001$) groups, and the Sometimes group scored significantly lower than the Always group ($p < .01$).

For the third MANOVA, the PAS Negative Affect subscale scores at Times 1 through 5 served as the dependent variables. There was a significant main effect for feigning group (F (2, 219) = 29.35, $p < .001$, partial $\eta^2 = .21$), but the main effect for time point was not significant (F (4, 216) = 1.76, $p = .14$, Wilk's $\Lambda = 0.97$, partial $\eta^2 = .03$). In this analysis, the interaction effect was statistically significant (F (8, 432) = 2.51, $p = .01$, Wilk's $\Lambda = 0.91$, partial $\eta^2 = .04$). This interaction effect can be seen visually in Figure 1. For the feigning group main effect, a Tukey post hoc test revealed that the mean PAS Total Scores in the Never group were significantly lower than both the Sometimes ($p < .001$) and Always ($p < .001$) groups, and the Sometimes group scored significantly lower than the Always group ($p = .001$). To better understand the significant interaction effect, independent-samples t-tests were conducted comparing each feigning group mean PAS Negative Affect score at each time point. As can be seen in Table 25, all of these t-tests were significant except for the Sometimes-Always comparison at Time 1. Of note, when a Bonferroni correction was applied to counteract the problem of multiple comparisons, the overall p-value of .05 became .004; the same comparisons remained significant. Cohen's *d* for the significant comparisons ranged from 0.46 to 1.34, indicating medium to large effect sizes.

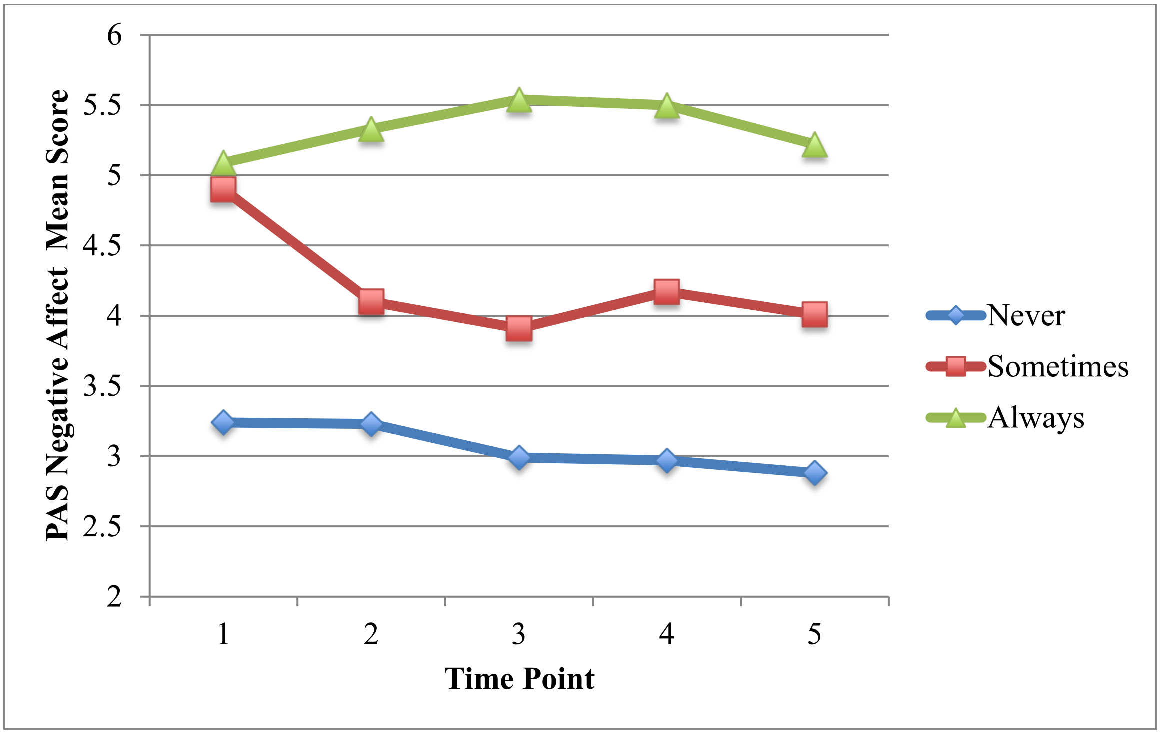

Figure 1. Interaction of Feigning Group and Time Point on PAS Negative Affect Scores

Table 25

Independent Samples T-Tests for PAS Negative Affect

Test	Time Point	Groups	n	*M*	*SD*	*t*	*df*	*p*	*d*
1	1	Never Sometimes	111 98	3.27 5.00	2.04 2.23	-5.87	207	< .001***	-0.84
2	1	Never Always	111 60	3.27 5.38	2.04 2.32	-5.88	169	< .001***	-0.99
3	1	Sometimes Always	98 60	5.00 5.28	2.23 2.32	-0.76	156	.445	-0.12
4	2	Never Sometimes	104 94	3.24 4.18	2.04 2.08	-3.21	196	.002**	-0.46
5	2	Never Always	104 58	3.24 5.43	2.04 2.30	-6.25	160	< .001***	-1.03
6	2	Sometimes Always	94 58	4.18 5.43	2.08 2.30	-3.45	150	.001**	-0.58
7	3	Never Sometimes	99 95	2.97 4.07	1.88 2.05	-3.91	192	< .001***	-0.56
8	3	Never Always	99 55	2.97 5.64	1.88 2.22	-7.91	152	< .001***	-1.33
9	3	Sometimes Always	95 55	4.07 5.64	2.05 2.22	-4.36	148	< .001***	-0.74
10	4	Never Sometimes	95 93	2.94 4.27	1.92 2.17	-4.46	186	< .001***	-0.65

11	4	Never Always	95 51	2.94 5.65	1.92 2.20	-7.73	144	< .001***	-1.34
12	4	Sometimes Always	93 51	4.27 5.65	2.17 2.20	-3.63	142	< .001***	-0.63
13	5	Never Sometimes	95 92	2.77 4.13	1.79 2.10	-4.78	185	< .001***	-0.70
14	5	Never Always	95 49	2.77 5.33	1.79 2.36	-7.27	142	< .001***	-1.28
15	5	Sometimes Always	92 49	4.13 5.33	2.10 2.36	-3.08	139	.002**	-0.55

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Bonferroni correction requires .004 for statistical significance.

For the fourth MANOVA, the BPRS Total Scores at Times 1, 3, and 5 served as the dependent variables. There were significant main effects for time point (F (2, 185) = 10.28, $p < .001$, Wilk's $\Lambda = 0.90$, partial $\eta^2 = .10$) and feigning group (F (2, 186) = 27.07, $p < .001$, partial $\eta^2 = .23$). The interaction effect was not significant (F (4, 370) = 2.12, $p = .08$, Wilk's $\Lambda = 0.96$, partial $\eta^2 = .02$). For the time point main effect, pairwise comparisons indicated that there were significant differences between Time 1 ($M = 34.24$, $SD = 8.57$) and Time 3 ($M = 31.69$, $SD = 6.90$, $p < .01$) and between Time 1 and Time 5 ($M = 31.14$, $SD = 6.37$, $p < .001$), but not between Time 3 and Time 5 ($p = .39$). For the feigning group main effect, a Tukey post hoc test revealed that the mean PAS Total Scores in the Never group were significantly lower than both the Sometimes ($p < .001$) and Always ($p < .001$) groups, and the Sometimes group scored significantly lower than the Always group ($p < .01$).

For the fifth MANOVA, the BPRS Anxious-Depressed composite scores at Times 1, 3, and 5 served as the dependent variables. There were significant main effects for time point (F (2, 184) = 6.27, $p < .01$, Wilk's $\Lambda = 0.94$, partial $\eta^2 = .06$) and feigning group (F (2, 185) = 21.74, $p < .001$, partial $\eta^2 = .19$). The interaction effect was not significant (F (4, 368) = 0.52, $p = .72$, Wilk's $\Lambda = 0.99$, partial $\eta^2 = .01$). For the time point main effect, pairwise comparisons indicated that there were significant differences between Time 1 ($M = 8.95$, $SD = 3.15$) and Time 3 ($M = 8.13$, $SD = 2.70$, $p < .01$) and between Time 1 and Time 5 ($M = 8.08$, $SD = 2.93$, $p < .01$), but not between Time 3 and Time 5 ($p = 1.00$). For the feigning group main effect, a Tukey post hoc test revealed that the mean PAS Total Scores in the Never group were significantly lower than both the Sometimes ($p < .01$) and Always ($p < .001$) groups, and the Sometimes group scored significantly lower than the Always group ($p = .001$).

For the sixth MANOVA, the BPRS Thought Disorders composite scores at Times 1, 3, and 5 served as the dependent variables. There were significant main effects for time point (F (2, 184) = 6.24, $p < .01$, Wilk's $\Lambda = 0.94$, partial $\eta^2 = .06$) and feigning group (F (2, 185) = 17.54, $p < .001$, partial $\eta^2 = .16$). The interaction effect was also statistically significant (F (4, 368) = 3.73, $p < .01$, Wilk's $\Lambda = 0.92$, partial $\eta^2 = .04$). This interaction effect can be seen visually in Figure 2, in which the mean scores for the three feigning groups become closer over time. For the time point main effect, pairwise comparisons indicated that there were significant differences between Time 1 ($M = 6.59$, $SD = 2.64$) and Time 3 ($M = 5.91$, $SD = 1.58$, $p < .01$) and between Time 1 and Time 5 ($M = 5.93$, $SD = 1.87$, $p < .01$), but not between Time 3 and Time 5 ($p = 1.00$). For the feigning group main effect, a Tukey post hoc test revealed that the mean PAS Total Scores in the Never group were significantly lower than both the Sometimes ($p = .001$) and Always ($p < .001$) groups, and the Sometimes group scored significantly lower than the Always group ($p = .017$). To better understand the significant interaction effect, independent-samples t-tests were conducted comparing each feigning group mean BPRS Thought Disorders score at each time point. As can be seen in Table 26, most of these t-tests were significant, with only the Time 1 Sometimes-Always, Time 3 Never-Sometimes, and Time 5 Sometimes-Always comparisons being non-significant. Of note, when a Bonferroni correction was applied to counteract the problem of multiple comparisons, the overall p-value of .05 became .006; the Time 5 Never-Sometimes comparison became non-significant. Cohen's *d* for the significant comparisons ranged from 0.42 to 0.98, with most effect sizes falling in the medium range.

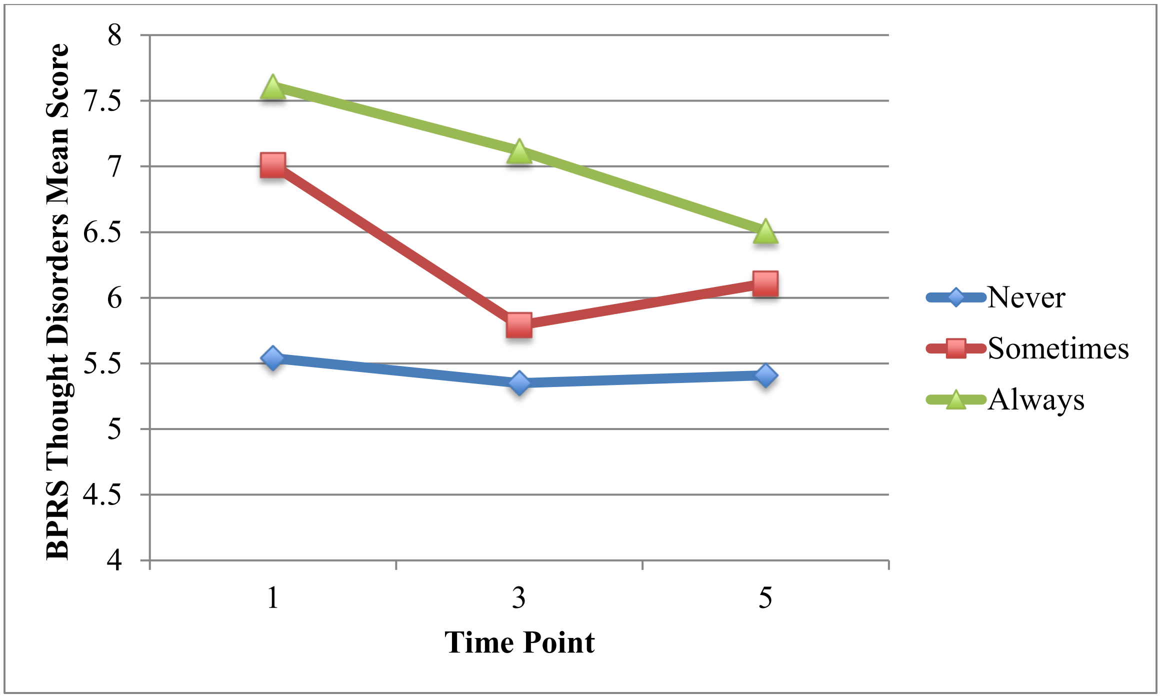

Figure 2. Interaction of Feigning Group and Time Point on BPRS Thought Disorders Scores

Table 26

Independent Samples T-Tests for BPRS Thought Disorders

Test	Time Point	Groups	n	*M*	*SD*	*t*	*df*	*p*	*d*
1	1	Never	104	5.61	1.46	-3.56	198	< .001***	-0.50
		Sometimes	96	6.74	2.87				
2	1	Never	104	5.61	1.46	-5.17	160	< .001***	-0.84
		Always	58	7.24	2.58				
3	1	Sometimes	96	6.74	2.87	-1.09	152	.277	-0.18
		Always	58	7.24	2.58				
4	3	Never	92	5.40	1.15	-1.93	181	.056	-0.29
		Sometimes	91	5.76	1.34				
5	3	Never	92	5.40	1.15	-5.56	140	< .001***	-0.98

		Always	50	6.86	1.98				
6	3	Sometimes	91	5.76	1.34	-3.92	139	< .001***	-0.69
		Always	50	6.86	1.98				
7	5	Never	79	5.37	1.25	-2.72	160	.007**	-0.42
		Sometimes	83	6.14	2.23				
8	5	Never	79	5.37	1.25	-3.92	123	< .001***	-0.73
		Always	46	6.48	1.92				
9	5	Sometimes	83	6.14	2.23	-0.86	127	.394	-0.16
		Always	46	6.48	1.92				

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Bonferroni correction requires .006 for statistical significance.

Research Question 3a

I next examined the extent to which changes in SIMS scores could be explained by corresponding changes in self-reported or clinician-rated psychopathology. To maximize sample size and statistical power, I focused on changes from Time 1 to Time 3, the first follow-up time point for which data was available for all three measures. First, I used three separate bivariate linear regression models to calculate residualized change scores for the SIMS, PAS, and BPRS Total Scores. Each model used the Time 1 score as the predictor and the Time 3 score as the outcome variable. These residual values were saved, as they represented the proportion of the follow-up score that was not linearly related to the baseline score. I then used a single multiple regression model to examine whether changes in SIMS scores were significantly correlated with the changes in PAS or BPRS scores. For this model, the residualized change scores for the PAS and BPRS served as predictors, and the residualized change scores for the SIMS served as the outcome variable.

Results indicated the PAS and BPRS residualized change scores significantly predicted the SIMS residualized change scores, F (2, 217) = 36.92, $p < .001$, $R^2 = .254$. This R^2 value indicates that 25.4% of the variance in the model was explained by the two predictors. Individually, both the PAS residualized change scores (B = .542, β = .470, t = 7.92, p < .001, r_{sp} = .464) and the BPRS residualized change scores (B = .161, β = .077, t = 2.09, p = .04, r_{sp} = .122) were significant predictors of SIMS residualized change scores. The squared semi-partial correlations for each of these predictors (r^2_{sp} = .215 and r^2_{sp} = .015, respectively) indicate that 21.5% of the variance in SIMS residualized change scores can be explained uniquely by the PAS residualized changes scores, and 1.5% can be explained uniquely by the BPRS residualized change scores.

Research Question 3b

I next examined the extent to which changes in SIMS scores could be explained by other participant characteristics, including housing placement (i.e., administrative segregation, general population, and San Carlos Correctional Facility), and mental health needs (i.e., needs versus no needs). To maximize sample size and statistical power, I focused on changes from Time 1 to Time 2. I first used a bivariate linear regression to calculate residualized change scores for the SIMS Total Score, using the Time 1 score as the predictor and the Time 2 score as the outcome variable. These residual values were saved, as they represented the proportion of the follow-up score that was not linearly related to the baseline score. I then used hierarchical regression analyses to predict SIMS residualized change scores using housing placement and mental health needs alternatively in the two steps.

In the first model, housing placement (represented by two dummy variables) served as the predictor in step 1 and mental health needs (dichotomous) served as the predictor in step 2. In the first step, housing placement did not significantly predict the SIMS residualized change scores, F (2, 258) = 1.10, p = .33, R^2 = .008. When mental health needs was added, the model remained non-significant, F (3, 257) = 1.12, p = .34, R^2 = .013.

In the second model, the predictors were swapped, with mental health needs (dichotomous) being entered in step 1 and housing placement (represented by two dummy variables) in step 2. In the first step, mental health needs did not significantly predict the SIMS residualized changed scores, F (1, 259) = 2.63, p = .11, R^2 = .010. When housing placement was added, the model again remained non-significant, F (3, 257) = 1.12, p = .34, R^2 = .013.

In summary, regardless of whether housing placement or mental health needs was added first, these participant characteristics did not significantly explain changes in SIMS scores from Time 1 to Time 2.

Research Question 3c

Finally, I examined whether changes in SIMS scores might be explained by regression toward the mean. In order to investigate this question, I computed the Pearson correlation of the Time 1-Time 2 score difference (absolute value) and the Time 1- mean Time 1 score across the sample. Results indicated a moderate-strength positive correlation between |Time 1-Time 2| (M = 5.61, SD = 6.27) and Time 1- mean Time 1 (M = 0.00, SD = 10.15), r = .40, p < .001, n = 261.

A Galton squeeze diagram was constructed (see Figure 3) to provide a visual aid of this regression toward the mean. In this diagram, data points falling at 1 on the X-axis indicate SIMS Total Scores at Time 1, and data points falling at 2 on the X-axis indicate the corresponding mean SIMS Total Scores at Time 2 for each Time 1 score. As suggested by the positive correlation between Time 1-Time 2 and Time 1- mean Time 1, there does appear to be a moderate regression toward the mean, in which the more extreme scores at Time 1 tend to become less extreme upon retesting.

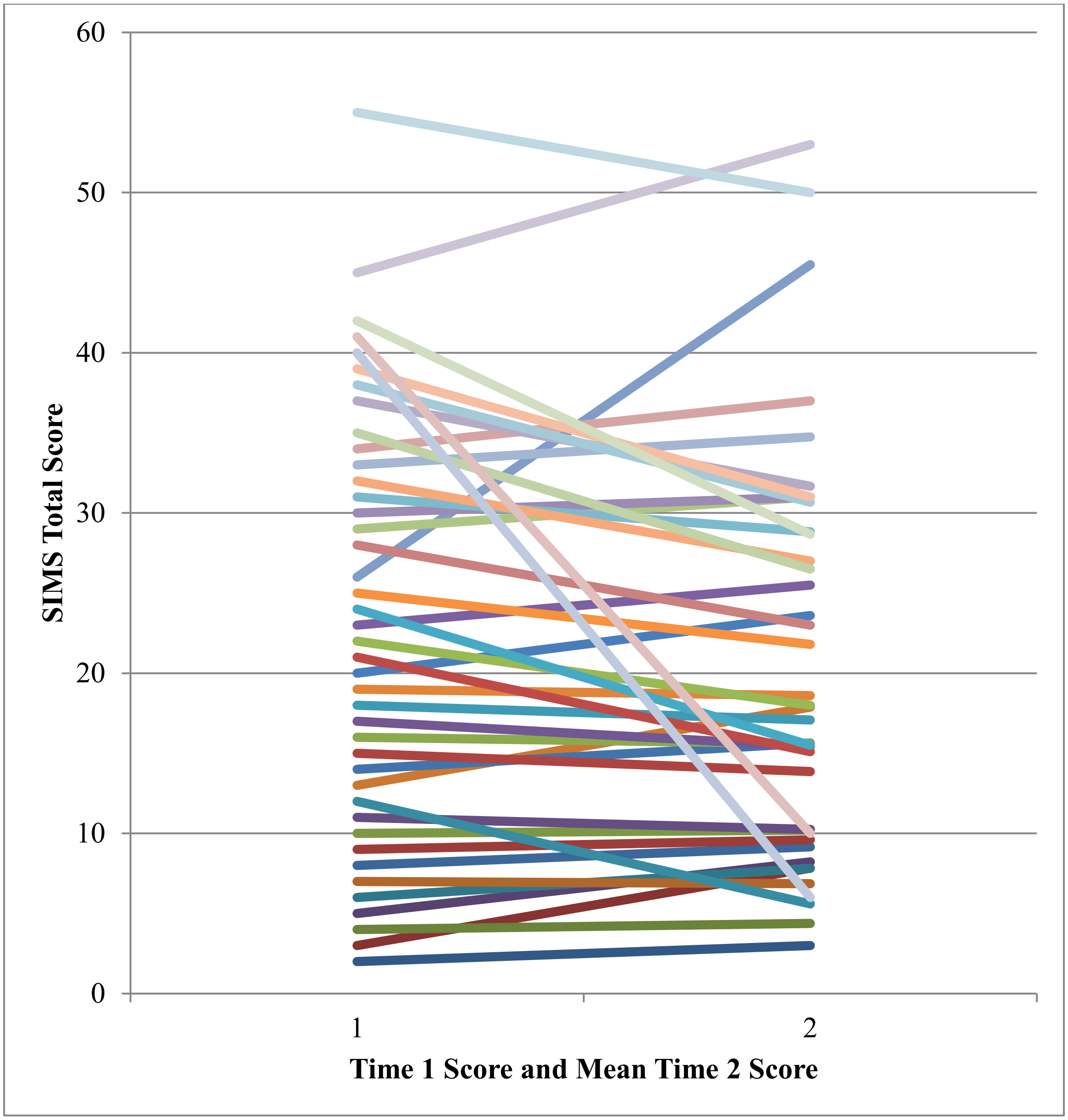

Figure 3. Galton Squeeze Diagram Showing Regression to the Mean from Time 1 to Time 2

CHAPTER IV

DISCUSSION

The primary purposes of this dissertation were to investigate consistency and patterns of feigning on a self-report screening measure, and to investigate whether changes in scores on this measure may be attributed to changes in scores on measures of psychopathology and/or to participant characteristics. A number of assessment tools have been developed for the detection of feigned psychopathology, and an extensive body of literature has documented strong support for the ability of these measures to detect feigning. However, most studies have investigated a single administration of each instrument, leaving a gap in our knowledge about the psychometric properties of repeated administrations and about how consistent examinees tend to be in their response styles over time. This dissertation used a publicly available dataset from a large study of inmates in Colorado (O'Keefe & Klebe, 2014), during which participants were administered numerous psychological assessments at five time points over a 12-month period. For this study, their patterns of scores on the Structured Interview of Malingered Symptomatology (SIMS, Widows & Smith, 2005) were evaluated across five administrations in order to explore the patterns of self-reported symptoms over time.

An additional research focus of this dissertation, which yielded particularly interesting results, was to compare the utility of two different cut-scores in examinees with and without genuine mental illness. Research on the SIMS as a screening measure of feigning has generally been positive, with very large effect sizes and good NPP and sensitivity estimates. However, a number of researchers have raised concerns regarding the potential for high rates of false positives among examinees with genuine symptoms.

While the SIMS manual recommends a Total Score cut-score of > 14, Rogers, Robinson, and Gillard (2014) recommended a significantly higher cut-score of > 44 in order to improve specificity of the measure. This dissertation investigated both recommended cut-scores, comparing rates of feigning among inmates with and without existing mental health needs.

Consistency in SIMS Scores and Classifications Over Time

To investigate consistency in SIMS scores over time, I first examined Pearson correlations between SIMS Total Scores and scale scores at all combinations of time points. For the Total Score, correlations ranged from $r = .61$ to $r = .83$, suggesting moderate to high reliability. Most SIMS scales were comparable in reliability, with relative weaknesses in test-retest reliability for the Affective Disorders and Symptom Combinations scales.

I used modified classification accuracy statistics (i.e., sensitivity, specificity, positive predictive power (PPP), negative predictive power (NPP), and overall accuracy) to examine the consistency of SIMS classifications across time. These statistics were modified such that instead of using a test to predict a known group, one SIMS classification was used to predict another SIMS classification at a later time point. When the Time 1/baseline classification was used to predict classifications at the other four time points, accuracy rates ranged from 75.00% to 80.40%. When the classification at each time point was used to predict the classification at the subsequent time point, accuracy rates ranged from 79.69% to 84.34%. As would be expected given the relative ease of predicting scores closer in time rather than farther apart; sensitivity, specificity, PPP, and NPP were all higher when predicting classifications using consecutive time points compared to using the baseline classification to predict all other time points. McNemar's tests were used to compare proportions of participants who scored in the feigning range at various combinations of time points. With the exception of the Time 1 – Time 5 comparison, all results were nonsignificant, suggesting that the proportions of participants who score in the feigning range at two different time points are more likely to be significantly different when the time points are farther apart.

Using the manual-recommended cut-score of > 14, inmates were sorted into one of three feigning subgroups (Never Feign, Sometimes Feign, Always Feign) based on the proportion of times they scored in the feigning range across administrations. I used six mix-model repeated measures MANOVAs to compare these three subgroups on measures of psychopathology, including the Personality Assessment Screener (PAS, Morey, 1997) and Brief Psychiatric Rating Scale (BPRS, Ventura et al., 1993). All six MANOVAs revealed significant main effects for feigning group, such that scores for participants in the Never group were significantly lower than the Sometimes and Always groups, and scores in the Sometimes group were significantly lower than the Always group. Thus, inmates whose response styles on the SIMS were sometimes or always classified as feigning tended to receive higher mean scores on both self-report and clinician-rated measures of psychopathology. Significant main effects for time point were found for the PAS Total Score, the BPRS Total Score, the BPRS Anxious-Depressed composite score, and the BPRS Thought Disorders composite score, such that scores tended to decrease from the baseline administration to later time points. Finally, interaction effects were found between feigning group and time point for the PAS Negative Affect clinical problem domain and the BPRS Thought Disorders composite. Independent samples t-tests were conducted comparing each feigning group score at each time point for these two subscales; most comparisons were significant with medium to large effect sizes for the PAS Negative Affect and medium effect sizes for the BPRS Thought Disorders.

Other Factors Corresponding with Changes in SIMS Scores

Despite moderate to high SIMS reliability over time, participants' scores tended to vacillate in and out of the feigning range. We examined several factors that could correspond with this inconsistency, including changes in scores on measures of psychopathology, participant characteristics, and regression to the mean.

I first examined the extent to which changes in SIMS scores could be explained by corresponding changes in self-reported or clinician-rated psychopathology. Bivariate linear regression models were used to calculate residualized changes scores for the SIMS, PAS, and BPRS Total Scores from Time 1 to Time 3 (i.e., 6-month follow-up), and these residuals were used in a single multiple regression model to examine whether changes in SIMS scores were significantly correlated with the changes in PAS or BPRS scores. Results indicated the PAS and BPRS residualized change scores significantly predicted the SIMS residualized change scores, together explaining 25.4% of the variance. Looking individually at each predictor, the PAS held more predictive weight, with 21.5% of the SIMS change variance explained uniquely by the PAS and 1.5% explained uniquely by the BPRS. Together, this information suggests that changes in SIMS are better predicted by a self-report measure of psychopathology than a clinician-rated measure. Thus, self-report measures may be more susceptible to response style changes compared to clinician-rated measures.

I next examined the extent to which changes in SIMS scores could be explained by other participant characteristics, including housing placement (i.e., administrative segregation, general population, and San Carlos Correctional Facility), and mental health needs (i.e., needs versus no needs). A bivariate linear regression model was used to calculate residualized change scores for the SIMS Total Score from Time 1 to Time 2 (i.e., 3-month follow-up), and then hierarchical regression analyses were used to predict these residuals using housing placement and mental health needs alternatively in the two steps. Regardless of whether housing placement or mental health needs was added first, these participant characteristics did not significantly explain changes in SIMS scores.

Next, I examined whether changes in SIMS scores might be explained by regression toward the mean, such that more extreme scores are closer to the mean upon reassessment. A Pearson correlation of the Time 1-Time 2 score difference (absolute value) and the Time 1- mean Time 1 score revealed a moderate-strength positive correlation. A Galton squeeze diagram was constructed to provide a visual aid of this moderate regression toward the mean, demonstrating that more extreme SIMS Total Scores at Time 1 tend to become less extreme upon retesting. This suggests that as participants complete additional administrations of the same feigning screener, extremely high scorers tend to endorse fewer items and extremely low scorers tend to endorse more items.

SIMS Cut-Score Comparisons

My final research question involved comparison of two cut-scores: the SIMS manual recommended cut-score of > 14 and the Rogers and colleagues (2014) recommended cut-score of > 44. Using a cut-score of > 14, 34.2% to 44.4% of participants scored in the feigning range at each time point. Using a cut-score of > 44, 0.7% to 3.1% of participants scoring in the feigning range at each time point. Over 58% of the sample was classified as feigning at least once using the cut-score of > 14, whereas less than 5% of the sample was classified as feigning at least once using the cut-score of > 44.

I next investigated the two recommended cut-scores separately for individuals with and without mental health needs documented by the prison. Using a cut-score of > 14, 75.5% of inmates with mental illness scored in the feigning range at least once (34.2% Always) whereas for inmates without mental illness, 35.7% scored in the feigning range at least once (6.1% Always). The odds of a participant with mental illness scoring in the feigning range at least once were over five times the odds of a participant without mental illness. Using a cut-score of > 44, only 6.4% of inmates with mental illness scored in the feigning range at least once (0.6% Always), and for participants without mental illness, only 1.7% scored in the feigning range at least once (0.0% Always). Due to the small number of participants who scored in the feigning range using the higher cut score, this odds-ratio was non-significant.

Finally, I used independent samples t-tests to compare mean SIMS Total Scores for inmates with and without mental illness. At all time points, inmates with mental illness received significantly higher mean Total Scores compared to inmates without mental illness. Mean scores for the no mental illness group ranged from 9.49 to 10.23; mean scores for the mental illness group ranged from 17.79 to 19.94. Pearson correlations between SIMS Total Scores and PAS/BPRS Total Scores revealed weak to moderate-strength positive correlations at all time points. Together, this information questions the

SIMS' ability to differentiate between genuine psychopathology and feigning response styles. These results are consistent with prior research findings of low SIMS specificity among individuals with genuine mental illness.

The correlations between feigning scores and scores on other measures of psychopathology appear to be an under-researched area of the literature. The SIMS has been found to correlate moderately with Symptom Checklist-90-Revised global severity index score, with $r = .52$ for the SIMS Total and r ranging from .28 to .49 for the subscales (Edens et al., 1999). These correlations are consistent with the correlations found in the current study. Regarding other measures of feigned psychopathology, only one study could be found reporting correlations with measures of psychopathology. In a sample of compensation-seeking veterans (Freeman, Powell, & Kimbrell, 2008), significant correlations were found between SIRS Total Scores and scores on the Clinician Administered PTSD Scale (CAPS-2) and the Diagnostic Interview for Borderlines (DIB), suggesting relations between the SIRS and reports of PTSD and borderline personality symptoms. For the CAPS-2 total and subscale scores, correlations with the SIRS ranged from $r = .33$ to $r = .45$, while the correlation for the DIB total score was $r = .51$. Correlations were non-significant between SIRS scores and Combat Exposure Scale scores, Beck Depression Inventory scores, or self-reported substance use. While the correlation between the SIMS and measures of psychopathology has been found concerning in both the current dissertation and at least one prior study, the dearth of research on other feigning measures does not allow for comparisons at this time. Future research should examine whether other feigning measures and screeners exhibit this same weakness in specificity.

Limitations and Strengths

There were several limitations to this study that must be acknowledged. The first, and largest, limitation was the lack of known groups for feigning status. While the available data allowed us to investigate consistency of SIMS scores over time, to compare various subgroups, and to investigate correspondence with other measures, there was no way to know whether each participant was truly feigning or responding honestly. Thus, while I found associations between changes in SIMS scores and changes on measures of psychopathology, it is unknown whether the SIMS may falsely determine genuine symptoms to be evidence of feigning or whether participants who feigned on the SIMS also tended to endorse more symptoms on the PAS and BPRS. It would benefit future research to investigate these research questions using known-groups samples. Second, all participants were adult males with adequate reading abilities who were incarcerated in the Colorado Department of Corrections, so it is unknown whether results could be generalized to females, juveniles or the elderly, inmates unable to read and write, prisons in other states. Third, this study used only the SIMS to assess response style, so generalizations to other feigning measures and screening measures are unknown.

Notwithstanding these limitations, this study had several strengths. First, this was the first study to examine patterns of scores of a feigning screener with five administrations, as previous research included only two administrations. Thus, my results provide an important contribution to the field's understanding of how scores on the SIMS may change over more than two administrations. Second, this was a relatively diverse sample, with approximately 60% of the sample reporting minority racial and ethnic backgrounds and with ages ranging from 17 to 59. Third, psychopathology was assessed using multiple methods, including both a self-report measure and a clinician-rated instrument.

Conclusions

These results support concerns raised by prior researchers regarding the potential for false positives among examinees with genuine mental illness. The manual-recommended cut-score of > 14 appears to be too low, as 75% of inmates with mental illness scored in the feigning range at least once, despite having no known reason to feign symptoms. If the majority of individuals with genuine mental illness score above the feigning cut-score, the screener may actually be capturing changes in symptomatology rather than changes in response style. At the same time, the > 44 cut-score recommended by Rogers and colleagues is likely to be too high, thus failing to capture all feigners. Perhaps a medial cut-score can be found that is less sensitive to changes in genuine psychopathology (i.e., improving specificity) but is still sensitive to changes in feigning (i.e., maintaining sensitivity). Van Impelen and colleagues (2014) suggested a range of cut-scores (i.e., > 16, > 19, 17 to 19, and > 24) from which examiners could choose, depending on whether the SIMS is used as a screener, as part of a larger test battery for feigning, and for detection of possible versus conclusive feigning. This is certainly a direction for future research. However, with either the > 14 or > 44 cut-scores, this feigning screener is unlikely to fulfill its very purpose. Thus, it may benefit forensic examiners to use other measures to screen and assess for feigned psychopathology until a more helpful cut-score is found for the SIMS.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Ben-Porath, Y. S., & Tellegen, A. (2008). *Minnesota Multiphasic Personality Inventory-2- Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Bender, S. D. (2018). Malingered traumatic brain injury. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 122-150). New York: Guilford.

Berry, D. T. R., & Schipper, L. J. (2007). Detection of feigned psychiatric symptoms during forensic neuropsychological examinations. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 226-263). New York, NY: Oxford University Press.

Boccaccini, M. T., & Hart, J. R. (2018). Response style on the Personality Assessment Inventory and other multiscale inventories. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 280-300). New York: Guilford.

Boccaccini, M. T., Murrie, D. C., & Duncan, S. A. (2006). Screening for malingering in a criminal-forensic sample with the Personality Assessment Inventory. *Psychological Assessment, 18*(4), 415-423. http://dx.doi.org/10.1037/1040-3590.18.4.415

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring, and interpretation* (Rev. ed.). Minneapolis, MN: University of Minnesota Press.

Burlingame, G. M., Seaman, S., Johnson, J. E., Whipple, J., Richardson, E., Chen, S., Earnshaw, D., Spencer, R., & Lehman, A. (2006). Sensitivity to change of the Brief Psychiatric Rating Scale- Extended (BPRS-E): An item and subscale analysis. *Psychological Services, 3*(2), 77-87. http://dx.doi.org/10.1037/1541-1559.3.2.77

Cima, M., Hollnack, S., Kremer, K., Knauer, E., Schellbach-Matties, R., Klein, B., & Merckelbach, H. (2003). Strukturierter Fragebogen Simulierter Symptome: Die Deutsche version des "Structured Inventory of Malingered Symptomatology: SIMS". Nervenarzt, 74, 977–986. http://dx.doi.org/10.1007/s00115-002-1438-5

Creech, S. K., Evardone, M., Braswell, L., & Hopwood, C. J. (2010). Validity of the Personality Assessment Screener in veterans referred for psychological testing. *Military Psychology, 22*, 465-473. http://dx.doi.org/10.1080/08995605.2010.513265

Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying persons motivated to malinger psychopathology. *Journal of the American Academy of Psychiatry and the Law, 27,* 387–396.

Edens, J. F., Penson, B. N., Smith, S. T., & Ruchensky, J. R. (2018). Examining the utility of the Personality Assessment Screener in three criminal justice samples. *Psychological Services.* Advance online publication. http://dx.doi.org/10.1037/ser0000251

Edens, J. F., Poythress, N. G., & Watkins-Clay, N. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment, 88*(1), 33-42. http://dx.doi.org/ 10.1080/00223890709336832

Frederick, R. I. (1997). Validity Indicator Profile manual. Minnetonka, MN: NCS Assessments.

Frederick, R. I. (2002) Review of the Validity Indicator Profile. *Journal of Forensic Neuropsychology, 2*(3-4), 125-145.

Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior, 24*(1), 59-82.

Freeman, T., Powell, M., Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research, 158*, 374-380.

Granacher, R. P., & Berry, D. T. R. (2018). Feigned medical presentations. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 243-253). New York: Guilford.

Green, D., & Rosenfeld, B. (2011). Evaluating the gold standard: A review and meta-analysis of the Structured Interview of Reported Symptoms. *Psychological Assessment, 23*(1), 95-107. http://dx.doi.org/10.1037/a0021149

Harris, M. R., & Resnick, P. J. (2003). Suspected malingering: Guidelines for clinicians. *Psychiatric Times, 20*(13), 68-71.

Hawes, S., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment, 21*, 112-124. http://dx.doi.org/10.1037/a0015036

Haydt, N. (2015). The DSM-5 and criminal defense: When does a diagnosis make a difference? *Utah Law Review, 2015*(4), 847-879.

Heilbronner, R. L., Sweet, J. J., Attix, D. K., Krull, K. R., Henry, G. K., & Hart, R. P. (2010). Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: The utility and challenges of repeat test administrations in clinical and forensic contexts. *The Clinical Neuropsychologist, 24*(8), 1267-1278. http://dx.doi.org/10.1080/13854046.2010.526785

Kelley, S. E., Edens, J. F., & Douglas, K. S. (2018). Concurrent validity of the Personality Assessment Screener in a large sample of offenders. *Law and Human Behavior, 42*(2), 156-166. http://dx.doi.org/10.1037/lhb0000276

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). New York, NY: Oxford University Press.

McLaughlin, J. L., & Kan, L. Y. (2014). Test usage in four common types of forensic mental health assessment. *Professional Psychology: Research and Practice, 45*(2), 128-135. http://dx.doi.org/10.1037/a0036318

Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C., Otto, R. K., Mossman, D., & Condie, L. O. (2018). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (4th ed.). New York, NY: The Guilford Press.

Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. Archives of Clinical Neuropsychology, 18, 145–152.

Millon, T., Grossman, S., & Millon, C. (2015). *Millon Clinical Multiaxial Inventory-IV (MCMI-IV) manual*. Bloomington, MN: NCS Pearson, Inc.

Millon, T., Millon, C., Davis, R., & Grossman, S. (2006). *Millon Clinical Multiaxial Inventory-III (MCMI-III) manual* (3rd ed.). Minneapolis, MN: NCS Pearson, Inc.

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (1997). *Personality Assessment Screener: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (2007). *Personality Assessment Inventory: Professional manual 2nd edition.* Odessa, FL: Psychological Assessment Resources.

O'Keefe, M. L. (2017). Reflections on Colorado's administrative segregation study. *NIJ Journal, 278*, 1-8. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/250346.pdf

O'Keefe, M. L., & Klebe, K. J. (2014). *Evaluation of the Psychological Effects of Administrative Segregation in Colorado, 2007-2010*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR31321.v1

O'Keefe, M. L., Klebe, K. J., Metzner, J., Dvoskin, J., Fellner, J., & Stucker, A. (2013). A Longitudinal study of administrative segregation. *Journal of the American Academy of Psychiatry and the Law, 41*(1), 49-60.

O'Keefe, M. L., Klebe, K. J., Stucker, A., Sturm, K., & Leggett, W. (2010). *One year longitudinal study of the psychological effects of administrative segregation*. Final report to the U.S. Department of Justice. Washington, DC: U.S. Department of Justice.

Overall, J. E., & Gorham, D. R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports, 10*, 799-812.

Overall, J. E., & Klett, C. J. (1972). *Applied Multivariate Analysis*. McGraw Hill, New York, NY.

Porcerelli, J. H., Hurrell, K., Cogan, R., Jeffries, K., & Markova, T. (2015). Personality Assessment Screener, childhood abuse, and adult partner violence in African American women using primary care. *Assessment, 22*(6), 749-752. http://dx.doi.org/10.1177/1073191115606213

Porcerelli, J. H., Kurtz, J. E., Cogan, R., Markova, T., & Mickens, L. (2012). Personality Assessment Screener in a primary care sample of low-income urban women. *Journal of Personality Assessment, 94*(3), 262-266. http://dx.doi.org/10.1080/00223891.2011.650304

Ray, C. L. (2009). The importance of using malingering screeners in forensic practice. *Journal of Forensic Psychology Practice, 9*, 138-146. http://dx.doi.org/10.1080/15228930802575508

Resnick, P. J. & Knoll, J. L. (2018). Malingered psychosis. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 98-121). New York, NY: The Guilford Press.

Resnick, P. J., West, S., & Wooley, C. N. (2018). The malingering of posttraumatic disorders. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 188-211). New York: Guilford.

Rogers, R. (2018a). An introduction to response styles. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 3-17). New York, NY: The Guilford Press.

Rogers, R. (2018b). Structured interviews and dissimulation. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 422-448). New York, NY: The Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Rogers, R., & Bender, S. D., (2003). Evaluation of malingering and deception. In A. Goldstein (Ed.), *Handbook of psychology: Forensic psychology* (Vol. 11, pp. 109–129): John Wiley & Sons, Inc.

Rogers, R., & Granacher, R. P. (2011). Conceptualization and assessment of malingering. In E. Y. Drogin, F. M. Dattilio, R. L. Sadoff, & T. G. Gutheil (Eds.), *Handbook of forensic assessment: Psychological and psychiatric perspectives* (pp. 659- 678). Hoboken, NJ: John Wiley & Sons, Inc.

Rogers, R., Robinson, E. V., & Gillard, N. D. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences and the Law, 32*, 455-466. http://dx.doi.org/10.1002/bsl.2131

Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms: Professional manual* (2nd ed.). Lutz, FL: Psychological Assessment Resources.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*(2), 160-177.

Rogers, R., Vitacco, M. J., & Kurus, S. J. (2010). Assessment of malingering with repeat forensic evaluations: Patient variability and possible misclassification on the SIRS and other feigning measures. *Journal of the American Academy of Psychiatry and the Law, 38*, 109-114.

Sellbom, M., & Bagby, R. M. (2010). Detection of overreported psychopathology with the MMPI-2-RF form validity scales. *Psychological Assessment, 22*(4), 757-767. http://dx.doi.org/10.1037/a0020825

Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI-2-RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment, 22*, 22-31. http://dx.doi.org/10.1037/a0018222

Smith, G. P. (2018) Brief screening measures for the detection of feigning and impression management. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of* *malingering and deception* (4^{th} ed., pp. 449-472). New York, NY: The Guilford Press.

Soliman, S., & Resnick, P. J. (2010). Feigning in adjudicative competence evaluations. *Behavioral Sciences and the Law, 28*, 614-629. http://dx.doi.org/10.1002/bsl.950

Tombaugh, T. N. (1996). Test of Memory Malingering (TOMM). Toronto, Canada: Multi-Health Systems.

Tombaugh, T. N. (2002). The Test of Memory Malingering (TOMM) in forensic psychology. *Journal of Forensic Neuropsychology, 2*(3-4), 69-96.

van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Interview of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist, 28*(8), 1336-1365. http://dx.doi.org/10.1080/13854046.2014.984763

Ventura, J., Lukoff, D., Nuechterlein, K. H., Liberman, R. P., Green, M., Shaner, A. (1993). Appendix 1: Brief Psychiatric Rating Scale (BPRS) Expanded Version (4.0) scales, anchor points, and administration manual. *International Journal of Methods in Psychiatric Research, 3,* 227-243.

Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology professional manual*. Odessa, FL: Psychological Assessment Resources.

Vitacco, M. J. (2018). Syndromes associated with deception. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4^{th} ed., pp. 83-97). New York, NY: The Guilford Press.

Wygant, D. B., & Lareau, C. R. (2015). Civil and criminal forensic psychological assessment: Similarities and unique challenges. *Psychological Injury and Law, 8*, 11-26. http://dx.doi.org/10.1007/s12207-015-9220-8

Wygant, D. B., Walls, B. D., Brothers, S. L., & Berry, D. T. R. (2018). Assessment of malingering and defensiveness on the MMPI-2 and MMPI-2-RF. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 257-279). New York, NY: The Guilford Press.

VITA
JESSICA R. HART

Education and Licensure

Ph.D. *Clinical Psychology,* Sam Houston State University (expected August 2020)
Emphasis in Forensic Psychology
Dissertation: Consistency of feigning measures scores over time: A study of SIMS scores across five administrations

LPA *Licensed Psychological Associate- Independent Practice*, No. 36738 (March 2014)
Texas State Board of Examiners of Psychologists

M.A. *Clinical and Counseling Psychology*, Midwestern State University (August 2013)
Thesis: Imposing cognitive load to facilitate lie detection: Using tapping as a secondary task

B.A. *Behavioral Science*, Concordia University- Irvine (May 2011)
Emphasis in Sociology, minor in Psychology
Magna Cum Laude, Honors Associate

Current Position

Aug 2019 – present *Predoctoral Intern*, Psychology Internship Program
Oregon State Hospital
Co-Training Directors: Ericia Leeper, Psy.D., and Kimberly McCollum, Psy.D.

Past Positions

Clinical

Aug 2016 – July 2019 *Student Forensic Evaluator*, Psychological Services Center
Sam Houston State University
Supervised by Mary Alice Conroy, Ph.D., Darryl Johnson, Ph.D., and Wendy Elliott, Ph.D.

June 2018 – June 2019 *Practicum Student,* Harris County Psychiatric Center
The University of Texas Health Science Center
Supervised by Alia Warner, Ph.D., and Elaheh Ashtari, Psy.D.

May 2017 – May 2018 *Practicum Student*, Walker County Department of Community Supervision and Corrections
Supervised by Wendy Elliott, Ph.D.

Aug 2016 – May 2017 *Practicum Student*, Psychological Services Center
Sam Houston State University
Supervised by Craig Henderson, Ph.D.

Mar 2014 – Jan 2017 *Contract Writer*, Forensic psychologist private practice

Supervised by Stacey Shipley, Psy.D.

Sept 2013 – July 2015 *Psychological Assistant*, North Texas State Hospital- Vernon Campus, Maximum Security Unit
Supervised by Jodi Blaszyk, Psy.D., and Michele Borynski, Ph.D.

Jan 2013 – May 2013 *Practicum Student*, North Texas State Hospital- Vernon campus, Maximum Security Unit
Supervised by Michele Borynski, Ph.D.

Apr 2012 – May 2013 *Practicum Student*, Psychology Clinic
Midwestern State University
Supervised by Laura Spiller, Ph.D., and Kelly Meyering, Psy.D.

Sept 2011 – Apr 2012 *Volunteer*, North Texas State Hospital- Wichita Falls campus

Research

Aug 2015 – July 2019 *Research Lab Member*, Boccaccini Lab
Sam Houston State University
Supervised by Marcus Boccaccini, Ph.D.

July 2016 – Nov 2018 *Research Lab Member,* Youth and Family Studies Lab
Sam Houston State University
Supervised by Amanda Venta, Ph.D.

July 2016 – Dec 2016 *Research Interviewer*, The LoneStar Project: Study of Offender Trajectories, Associations, and Reentry
University of Colorado-Boulder, Arizona State University, and Sam Houston State University
Funded by the National Institute of Justice
Bronze Tier Interviewer

Mar 2016 – Aug 2016 *Research Interviewer*, Early Markers of Suicidal Behaviors in Youth: A Cross Sectional Study
University of Texas Health Science Center at Houston and University of Houston

Jan 2012 – May 2013 *Research Lab Member*, Sexual Aggression Research Lab
Midwestern State University
Supervised by Laura Spiller, Ph.D.

Aug 2011 – May 2012 *Research Assistant*, Midwestern State University
Supervised by David Carlston, Ph.D.

Oct 2010 – May 2011 *Research Assistant*, Concordia University- Irvine

Supervised by Robert Flores de Apodaca, Ph.D.

Teaching

June 2019 — *Invited Presenter*, Mental State at the Time of the Offense: Basics of the Insanity Evaluation
UTHealth-HCPC PGY2 Resident Forensic Didactic Series

June 2019 — *Invited Presenter*, Client Suicide: What Next?
SHSU Clinical Psychology Ph.D. Program, Spring Speaker

Jan 2017 – May 2017 — *Teaching Assistant*, Personality Psychology (PSYC 4331)
Sam Houston State University

Aug 2016 – Dec 2016 — *Teaching Assistant*, Abnormal Psychology (PSYC 3331)
Sam Houston State University

May 2016 – Aug 2016 — *Teaching Assistant*, Developmental Psychology (PSYC 3374)
Sam Houston State University

Jan 2016 – May 2016 — *Teaching Assistant*, Developmental Psychology (PSYC 3374)
Sam Houston State University

Jan 2013 – May 2013 — *Teaching Assistant*, General Psychology (PSYC 1103)
Midwestern State University

Aug 2012 – Dec 2012 — *Teaching Assistant*, General Psychology (PSYC 1103)
Midwestern State University

Supervision

May 2017 – Aug 2017 — *Peer Supervisor*, Psychological Services Center
Beginning Doctoral Practicum

Jan 2017 – May 2017 — *Peer Supervisor,* Psychological Services Center
Theory and Research in Psychotherapy I

Committee

Aug 2017 – Aug 2019 — *Student Liaison*, Teaching, Training, and Careers Committee
American Psychology-Law Society (APA Div. 41)

Book Chapter

Boccaccini, M. T., & **Hart, J. R.** (2018). Response style on the Personality Assessment Inventory and other multiscale inventories. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 280-300). New York: Guilford.

Journal Article

Hart, J. R., Venta, A., & Sharp, C. (2017). Attachment and thought problems in an adolescent inpatient sample: The mediational role of theory of mind. *Comprehensive Psychiatry, 78*, 38-47. http://dx.doi.org/10.1016/j.comppsych.2017.07.002

Conference Presentations

Hart, J. R., & Thomas, K. (Under review). *Borderline traits and apparent overreporting: Exploring the relation between personality and response validity.* Paper submitted for presentation at the annual convention of the American Psychology-Law Society, New Orleans, LA.

Hart, J. R. & Boccaccini, M. T. (2019, March). *False positives for SIMS cut scores among inmates with and without mental illness.* Paper presented at the annual convention of the American Psychology- Law Society, Portland, OR.

Ball Cooper, E., **Hart, J.**, & Venta, A. (2019, March). *The moderating role of inmate housing placement on hopelessness and psychopathology.* Paper presented at the annual convention of the American Psychology-Law Society, Portland, OR.

Hart, J. R., Reinhard, E. E., Boccaccini, M. T., Domino, M., & Cooper, V. (2018, March). *Correspondence between Structured Interview of Reported Symptoms' (SIRS) scores and clinicians' opinions of malingering.* Paper presented at the annual convention of the American Psychology- Law Society, Memphis, TN.

Hart, J. R., Kavish, N., & Boccaccini, M. T. (2017, March). *Feigning in a correctional sample: Associations of SIMS scores with elevations on other measures.* Poster presented at the annual convention of the American Psychology- Law Society, Seattle, WA.

Abate, A., Harmon, J., Marshall, K., **Hart, J.,** Ball, E., Henderson, C., Desforges, D., & Venta, A. (2017, March). *Perceptions of the legal system and recidivism: Investigating the mediating role of perceptions of chances for success in juvenile offenders*. Paper presented at the annual convention of the American Psychology-Law Society, Seattle.

Hart, J. R., Sharp, C., & Venta, A. (2016, November). *Attachment and thought problems in an adolescent inpatient population: The mediational role of theory of mind*. Paper presented at the annual convention of the Texas Psychological Association, Austin, TX.

Magyar, M. S., Ball, E. M., & **Hart, J. R.** (2016, June). *Borderline features: Critical mediator in the relation between childhood maltreatment and diverse aggressive and delinquent features among justice-involved youth.* Paper presented at the annual convention of the International Association of Forensic Mental Health Services, New York, NY.

Hart, J. R., Magyar, M. S., Ball, E. M., Camins, J., Ridge, B., & Edens, J. (2016, March). *Using the Personality Assessment Inventory-Adolescent to predict high-risk behaviors among juvenile male offenders.* Paper presented at the annual convention of the American Psychology- Law Society, Atlanta, GA.

Abate, A. C., Magyar, M., Ball, E., Ricardo, M., **Hart, J.**, & Edens, J. (2016, March). *Use of the Personality Assessment Inventory-Adolescent to assess trauma-related symptoms in justice-involved youth*. Paper presented at the annual convention of the American Psychology- Law Society, Atlanta, GA.

Cook, J. R., Carlston, D. L., Callaway, L. R., & Price, A. (2013, April). *Imposing cognitive load to facilitate lie detection: Tapping as a secondary task*. Poster presented at the annual convention of the Southwestern Psychological Association, Fort Worth, TX.

Strug, A. M., **Cook, J. R.**, Spiller, L. C., & Davis, J. (2012, April). *Measuring college women's perception of competition for their romantic partner*. Paper presented at the annual convention of the Southwestern Psychological Association, Oklahoma City, OK.

Conference Reviewer Positions

2018	*Student Reviewer*, American Psychology-Law Society (APA Div. 41)
2017	*Student Reviewer*, American Psychology-Law Society (APA Div. 41)
2016	*Student Reviewer*, American Psychology-Law Society (APA Div. 41)
2015	*Student Reviewer*, American Psychology-Law Society (APA Div. 41)
2014	*Student Reviewer*, American Psychology-Law Society (APA Div. 41)

Certifications

2017	Teaching Assistant Certification Series, Sam Houston State University
2016	Columbia Suicide Severity Rating Scale (C-SSRS)

Professional Affiliations

Sep 2016 – present	Texas Psychological Association
Apr 2012 – present	American Psychology- Law Society (APA Div. 41)
Mar 2012 – present	American Psychological Association
Jan 2012 – Jan 2014	Southwestern Psychological Association
May 2010- present	Psi Chi, National Honor Society in Psychology

Honors and Awards

2019	*Student Travel Award*, American Psychology-Law Society
2018 – 2019	*AP-LS Childcare Grant*, American Psychology-Law Society, Professional Development of Women Committee
2017	*2017 Continuing Education (CE) Pre-Conference Grant*, American Psychology-Law Society
2016	*SHAPA Travel Award*, Sam Houston Area Psychological Association
2014	*"Watch Me Shine" Annual Award*, North Texas State Hospital

2011 – 2013	*Graduate Merit Scholarship*, Midwestern State University
2011 – 2012	*Arbor Creek Scholarship*, Maven Management
2011 – 2012	*Chellgren Family Scholarship*, Omicron Delta Kappa Foundation
2008 – 2011	*Presidential Scholarship*, Concordia University
2010	*Leader of the Year Award,* Omicron Delta Kappa, Concordia University